

UNIVERSIDADE FEDERAL DO PARANÁ

VIVIANE DEMETRIO DO NASCIMENTO

ANÁLISE DO SEQUENCIAMENTO EM LARGA ESCALA DO GENOMA DE  
*Apareiodon* sp. NA CARACTERIZAÇÃO E LOCALIZAÇÃO *IN SITU* DE  
ELEMENTOS REPETITIVOS

CURITIBA

2019

VIVIANE DEMETRIO DO NASCIMENTO

ANÁLISE DO SEQUENCIAMENTO EM LARGA ESCALA DO GENOMA DE  
*Apareiodon* sp. NA CARACTERIZAÇÃO E LOCALIZAÇÃO *IN SITU* DE  
ELEMENTOS REPETITIVOS

Tese apresentada ao Programa de Pós-Graduação em Genética, Setor de Ciências Biológicas da Universidade Federal do Paraná (UFPR) como parte dos requisitos para obtenção de título de Doutora em Ciências Biológicas, área de concentração Genética.

Orientador: Prof. Dr. Marcelo Ricardo Vicari  
Coorientadora: Dra. Michelle Orane Schemberger

CURITIBA

2019

Universidade Federal do Paraná. Sistema de Bibliotecas.  
Biblioteca de Ciências Biológicas.  
(Dulce Maria Bieniara – CRB/9-931)

Nascimento, Viviane Demetrio do

Análise do sequenciamento em larga escala do genoma de *Apareiodon*  
*sp.* na caracterização e localização *in situ* de elementos repetitivos. /  
Viviane Demetrio do Nascimento. – Curitiba, 2019.  
165 p.: il.

Orientador: Marcelo Ricardo Vicari

Coorientadora: Michelle Orane Schemberger

Tese (doutorado) - Universidade Federal do Paraná, Setor de Ciências  
Biológicas. Programa de Pós-Graduação em Genética.

1. Sequenciamento de nucleotídeos em larga escala 2. Elementos de  
DNA transponíveis 3. DNA satélite 4. Cromossomos sexuais I. Título II.  
Vicari, Marcelo Ricardo III. Schemberger, Michelle Orane IV. Universidade  
Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-  
Graduação em Genética.

CDD (20. ed.) 574.87322



MINISTÉRIO DA EDUCAÇÃO  
SETOR SETOR DE CIÊNCIAS BIOLÓGICAS  
UNIVERSIDADE FEDERAL DO PARANÁ  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO GENÉTICA -  
40001016006P1

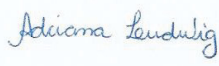
## TERMO DE APROVAÇÃO


Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GENÉTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **VIVIANE DEMETRIO DO NASCIMENTO** intitulada: **Análise do sequenciamento em larga escala do genoma de *Apareiodon sp.* na caracterização e localização *in situ* de elementos repetitivos**, após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

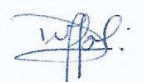
A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.


CURITIBA, 29 de Março de 2019.

  
MARCELO RICARDO VICARI  
Presidente da Banca Examinadora

  
ADRIANA LUDWIG  
Avaliador Externo (ICC)

  
CESAR MARTINS  
Avaliador Externo (UNESP/BOT)

  
DANIEL PACHECO BRUSCHI  
Avaliador Interno (UFPR)

  
KALINE ZIEMNICZAK  
Avaliador Externo (UNESPAR)



*Dedico este trabalho a todos que acreditam  
no poder da educação e do conhecimento  
como ferramentas na transformação do ser  
humano.*

## AGRADECIMENTOS

A **Deus** por ter me dado condições para lutar e fé para acreditar que tudo é possível.

Ao meu pai **Genésio**, que mesmo ausente neste momento da minha vida, conduziu meus passos até aqui com seus ensinamentos e valores. Sempre lhe manterei vivo dentro de mim. Saudades eternas!

A minha mãe **Gercina** pelo amor, orações e por ter entendido minhas ausências. Te amo!

Aos meus irmãos **Lígia, Wagner, Djonatas e Andrey**. Meus fiéis escudeiros. Obrigada por todo apoio, preocupação e amor incondicional. Em especial meu irmão **Wagner** por ter me acompanhado e contribuído com este trabalho. Amo vocês!

Agradecimento especial ao meu amor **Daniel**. As palavras são insuficientes para expressar, de forma fiel, toda a minha gratidão. Obrigada por caminhar comigo, por segurar a minha mão, por viver os meus sonhos, por acreditar em mim, pelo amor, pela paciência e companheirismo. Te amo!

Aos meus sobrinhos **Nathan, Vinícius, Lucas, Samuel, Livia** e meu afilhado **Giovane** que me permitem retornar a infância e partilhar momentos maravilhosos de brincadeiras e descontração.

A todos os meus familiares que me apoiaram, em especial a minha sogra **Regina**, minha tia **Maria Joana**, minha cunhada **Valérya** e minha comadre **Rafaela** que com pequenos detalhes contribuíram grandemente durante esta jornada.

Ao meu orientador **Prof. Dr. Marcelo Ricardo Vicari** pela orientação durante seis anos. Obrigada pelo voto de confiança, pela amizade, pelos ensinamentos e pela imensurável contribuição para minha formação. É admirável a sua inteligência, sua sabedoria e incansável dedicação à pesquisa.

A minha coorientadora **Dra. Michele Orane Schemberger** por toda ajuda, atenção, pelo conhecimento compartilhado e pela amizade. Você é uma grande pesquisadora. Te admiro e torço muito por você!!!!

A professora **Dra. Viviane Nogaroto Vicari** pelo auxílio, pela compreensão e pela amizade.

Aos Professores **Dr. Daniel Bruschi** e **Dra. Marta Margarete Cestari** por acompanharem o desenvolvimento deste trabalho ao longo desses anos. Obrigada pela disponibilidade e contribuições.

Aos professores **Dr. Cesar Martins**, **Dr. Orlando Moreira Filho** e **Dr. Rafael Bueno Noleto** pelo suporte e contribuições.

A minha grande amiga **Kaline** pela parceria de sempre, pelos conselhos, pela palavra de apoio com tanta serenidade. É sempre muito bom estar perto de você.

Ao meu grande amigo (irmão) **Lucas**. Para você vou fazer uso frase de Charles Chaplin "*Um dia sem risada é um dia desperdiçado*". Sou privilegiada em ter você como amigo. Obrigada pela partilha, pela amizade, pelas incontáveis risadas, pelo companheirismo, pelas loucuras e pelas guloseimas. Te adoro de coração!

Aos amigos do laboratório CBSF **Rafael**, **Caroline**, **Marcela** e **Felipe** pela amizade e acolhimento. Em especial a **Thaís**, pela amizade, pela força, incentivo, apoio incondicional. A **Geize** pela amizade, pelo incentivo, pela parceria nas viagens e no karaokê. Ao **Sebastião** pela amizade, pelas risadas e toda ajuda. A **Larissa** pela amizade e apoio nas atividades do laboratório. Ao **Matheus** pela amizade e por estar sempre disposto a contribuir.

A todos os membros da Pós-graduação em Genética da Universidade Federal do Paraná pelo incentivo e apoio.

A Universidade Estadual de Ponta Grossa e ao Laboratório de Biologia Cromossômica: Estrutura e Função pela infraestrutura.

Ao Instituto Brasileiro do Meio Ambiente (IBAMA/MMA/SISBIO nº: 15117) pela autorização para coleta do material utilizado.

As agências de fomento do CNPq, CAPES, SETI, Fundação Araucária e FAPESP.

Ao concluir estes agradecimentos encontro uma lacuna na minha gratidão ao lembrar de inúmeras pessoas que embora não tenham participado diretamente deste trabalho contribuíram grandemente para minha formação acadêmica e desenvolvimento como ser humano.

Assim, gostaria de agradecer a todos os meus professores, em especial a **Profª Marli Dancosky**, que me ensinou as primeiras letras a **Profª Maderli Copacheski Hampf** por ter despertado em mim o gosto pela Ciência e a **Profª Shirley Santos** que embora não tenha sido minha professora oficialmente foi uma grande incentivadora na minha vida acadêmica. Assim, deixo aqui o meu reconhecimento e minha mais profunda admiração a todos estes profissionais, que no exercício da mais nobre profissão, desempenharam um papel fundamental na minha trajetória.

Por fim, agradeço a todos os brasileiros que ao longo dos meus 28 anos de vida acadêmica, integralmente em instituições públicas, contribuíram para a efetivação da minha formação. Assim, em retribuição deixo aqui o meu compromisso, no papel de professora, de trabalhar incansavelmente pela educação do nosso país.

*O grande inimigo do conhecimento não é a ignorância,  
mas a ilusão do conhecimento.*

**Stephen Hawking**

*Hoje me sinto mais forte, mais feliz, quem sabe  
Só levo a certeza de que muito pouco sei ou nada sei.*

**Almir Sater e Renato Teixeira**



## RESUMO

Nas últimas décadas os avanços na Era Genômica têm impulsionado uma compreensão mais detalhada dos principais constituintes do material genético e possibilitado investigações com diferentes abordagens, principalmente no que se refere aos elementos repetitivos. Recentemente, um representante da família de peixes Parodontidae, portador de cromossomos sexuais ZW (*Apareiodon* sp.), teve o genoma sequenciado. A anotação evidenciou que a fração repetitiva compõe 36% do genoma de *Apareiodon* sp. A análise das sequências demonstrou uma grande diversidade de elementos, com 56 superfamílias de elementos transponíveis (TEs) e 20 tipos de expansões microssatélites, os quais ainda são poucos explorados e compreendidos quanto a sua estrutura e função. Este estudo teve por objetivo a caracterização molecular dos TEs Gypsy e En/Spm e sua distribuição no genoma de *Apareiodon* sp., com enfoque especial na evolução cariotípica e origem dos cromossomos sexuais em Parodontidae. Nossos dados evidenciaram altas taxas de substituição nucleotídica e diferentes níveis de degeneração das regiões regulatórias e codificadoras nas sequências de Gypsy e En/Spm, sugerindo a inatividade destes elementos no genoma de *Apareiodon* sp. Em outra perspectiva, nossos dados de mapeamento cromossômico evidenciaram um acúmulo do elemento En/Spm e microssatélites (TA)<sub>n</sub> e (GA)<sub>n</sub> colocalizados no cromossomo sexual W em Parodontidae. A análise dos *contigs* contendo En/Spm permitiu inferir que sequências deste TE atuaram na gênese de expansões microssatélites e de um DNA satélite presentes nos cromossomos sexuais. Para estas repetições *in tandem* foram atribuídas uma possível coparticipação no processo de heterocromatização do cromossomo W e, conseqüentemente, na diferenciação dos cromossomos sexuais heteromórficos em Parodontidae. Em contrapartida, o elemento Gypsy parece não ter atuado neste processo, sem acúmulos diferenciais no cromossomo W. Possivelmente, a invasão genômica de Gypsy é remota e sua inatividade é anterior a origem dos cromossomos sexuais heteromórficos em *Apareiodon* sp. Em adição, uma evidência de co-opção molecular foi encontrada para Gypsy, onde detectamos o domínio íntegro de PNMA (antígenos paraneoplásicos Ma) interno as suas sequências. PNMA é derivada de Gypsy, mas até o momento foi encontrada apenas em mamíferos, sugerindo uma possível atuação de Gypsy em outros mecanismos moleculares no genoma de *Apareiodon* sp. Nossos dados mostraram a inatividade dos elementos Gypsy e En/Spm no genoma de *Apareiodon* sp. Contudo, o conjunto dos dados indicam uma importante atuação destes elementos na diversidade genômica e evolução cariotípica de Parodontidae.

**Palavras-chave:** sequenciamento em larga escala, elemento transponível, co-opção molecular, cromossomo sexual heteromórfico W, DNA satélite.

## ABSTRACT

In the last decade the advances in the Genomic Age have impelled a more detailed understanding of the main constituents of the genetic material and made possible genomic investigations with different approaches, mainly with regard to the repetitive elements. Recently, a representative of the Parodontidae fish family, that presented ZW sex chromosomes (*Apareiodon* sp.), had the genome sequenced. The annotation showed that the repetitive fraction makes up 36% of the genome of *Apareiodon* sp. with wide diversity of elements (56 superfamilies of transposable elements (TEs) and 20 types of microsatellite expansions), and this sequences are still few explored and understood as to their structure and function. This study aimed at the molecular characterization of the Gypsy and En/Spm TEs and verify their distribution in the genome of *Apareiodon* sp., with a special focus on the karyotype evolution and origin of the sex chromosomes in Parodontidae. Our data evidenced high nucleotide substitution rates and different levels of degeneration of the regulatory and coding regions in the Gypsy and En/Spm sequences, suggesting the inactivity of these elements in the genome of *Apareiodon* sp. In another perspective, the chromosome mapping evidenced an accumulation of the En/Spm element and microsatellites (TA)<sub>n</sub> and (GA)<sub>n</sub> colocalized in the W sex chromosome in Parodontidae. The analysis of the En/Spm contigs contents allowed us to infer that sequences of this TE acted in the genesis of microsatellite expansions and of a satellite DNA present in the sex chromosomes. For these in-tandem repeats DNAs, a possible co-participation in the W sex chromosome heterochromatization and in the heteromorphic sex chromosomes differentiation into Parodontidae were attributed. On the other hand, the Gypsy element does not seem to have acted in the differential accumulations on the W sex chromosome. Possibly the genomic invasion of Gypsy is remote and its inactivity precedes the origin of the heteromorphic sex chromosomes in *Apareiodon* sp. In addition, evidence of molecular co-option was found for Gypsy, where we detected the complete domain of PNMA (paraneoplastic antigens Ma) in their sequences. PNMA is derived from Gypsy, but has so far been found only in mammals, suggesting a possible role of Gypsy in other molecular mechanisms in the genome of *Apareiodon* sp. Our data showed the inactivity of the Gypsy and En/Spm elements in the genome of *Apareiodon* sp. However, the data set indicate an important performance of these elements in the genomic diversity and karyotype evolution of Parodontidae.

**Key words:** Next-generation sequencing, transposable element, molecular co-option, W chromosome, satellite DNA.

## LISTA DE ILUSTRAÇÕES

<b>FIGURA 1.</b> Classificação dos elementos transponíveis.....	<b>27</b>
<b>FIGURA 2.</b> Representação esquemática do mecanismo de retrotransposição de um elemento LTR.....	<b>28</b>
<b>FIGURA 3.</b> Representação esquemática do mecanismo de transposição de transposons clássicos.....	<b>30</b>
<b>FIGURA 4.</b> Representantes da família Parodontidae.....	<b>39</b>
<b>FIGURA 5.</b> Caracteres morfológicos de exemplar do gênero <i>Apareiodon</i> .....	<b>44</b>
<b>FIGURA 6.</b> Pontos de coleta dos exemplares da família Parodontidae .....	<b>49</b>
<b>FIGURA 7.</b> Apresentação geral das etapas desenvolvidas durante este estudo.....	<b>51</b>

### CAPÍTULO 1

<b>Fig. 1</b> Landscape analysis of the RTE-Gypsy in <i>Apareiodon</i> sp. genome. The graphs show the Gypsy divergence from its consensus (x-axis) in relation to the number of copies on the genome (y-axis). Peaks represent insertion the number of Gypsy copies on the genome (y-axis). The peak represents insertion waves of Gypsy into the genome. In a) female and, b) male.....	<b>74</b>
<b>Fig. 2</b> Schematic representation of the possible combinations of the protein domains of Gypsy found in <i>Apareiodon</i> sp., in the 121 contigs analyzed. The numbering on the right indicates the number of times this setting was detected. Distances between domains, size, integrity, and location do not match the original sequences.....	<b>75</b>
<b>Fig. 3</b> Schematic representation of Gypsy LTRs, detected in Retroector (Sperber et al. 2007), showing the main conserved regions.....	<b>76</b>
<b>Fig. 4</b> Representation of possible types of ORFs detected in 121 contigs analyzed for Gypsy in <i>Apareiodon</i> sp. Sequences with 1, 2, 3 and 4 ORFs were founded.....	<b>77</b>
<b>Fig. 5</b> <i>In situ</i> location of the Gypsy RTE and karyotypes arrange in female individuals of Parodontidae species: <i>A. piracicabae</i> , <i>Apareiodon</i> sp. (a), <i>Parodon hilarii</i> (b) and; <i>Apareiodon affinis</i> (c). FISH using Gypsy probe (red) obtained from <i>Apareiodon</i> sp. (set of primers 3).....	<b>78</b>
<b>Supplementary Fig. 1A</b> Plotcon graphs demonstrating the similarity of the sequences for each GAG region domain (frame 1) and PNMA (frame 2) in <i>Apareiodon</i> sp. (in D, alignment using 12 PNMA of the <i>Apareiodon</i> sp. and; in E, alignment using integre PNMA of <i>Apareiodon</i> with PNMA of mammals).....	<b>85</b>
<b>Supplementary Fig. 1B</b> Plotcon graphs demonstrating the similarity of the sequences for each protein domain of POL in <i>Apareiodon</i> sp. ....	<b>86</b>
<b>Supplementary Fig. 1C</b> Plotcon graphs demonstrating the similarity of the 5' and 3' LTR sequences in <i>Apareiodon</i> sp.....	<b>87</b>
<b>Supplementary Fig. 2</b> Alignment MUSCLE of the intact Zf - CCHC sequences showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of <i>Apareiodon</i> sp. ....	<b>87</b>

<b>Supplementary Fig. 3</b> MUSCLE alignment of the SCAN intact sequences evidencing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of <i>Apareiodon</i> sp.....	<b>87</b>
<b>Supplementary Fig. 4</b> Alignment MUSCLE of the intact sequences of Aspartic Proteinase showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of <i>Apareiodon</i> sp.....	<b>88</b>
<b>Supplementary Fig. 5</b> Alignment MUSCLE of the Reverse transcriptase intact sequences showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of <i>Apareiodon</i> sp.....	<b>88</b>
<b>Supplementary Fig. 6</b> Alignment MUSCLE of the RNase H intact sequences showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of <i>Apareiodon</i> sp.....	<b>89</b>
<b>Supplementary Fig. 7</b> MUSCLE alignment of Integrase sequences of Integrase showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of <i>Apareiodon</i> sp.....	<b>90</b>
<b>Supplementary Fig. 8</b> Alignment MUSCLE of the intact sequences of the Chromo domain evidencing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of <i>Apareiodon</i> sp.....	<b>91</b>
<b>Supplementary Fig. 9</b> Alignment MUSCLE of the intact Zf - HHCC sequences showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of <i>Apareiodon</i> sp.....	<b>91</b>
<b>Supplementary Fig. 10</b> Alignment MUSCLE of the PNMA sequences showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of <i>Apareiodon</i> sp.....	<b>92</b>
<b>Supplementary Fig. 11</b> Agarosis electrophoresis gel 3% showing IN, RT and PNMA amplification of <i>Apareiodon</i> sp. using genomic DNA (left) and, cDNA of testis (right).....	<b>92</b>

## CAPÍTULO 2

- Fig. 1** Repeat Landscape do transposon En/Spm em *Apareiodon* sp. O gráfico mostra no eixo X a divergência para cada consenso em relação ao número de cópias no genoma, apresentada no eixo Y. a) macho b) fêmea.....114
- Fig. 2** Localização *in situ* do transposon En/Spm (vermelho) em cariótipos de fêmeas de espécies da família Parodontidae: Em (a) *A. piracicabae*, (b) *P. nasus*, (c) *Apareiodon* sp., (d) *P. hilarii* e, (e) *A. affinis*.....115
- Fig. 3** Localização *in situ* dos microssatélites (GA)<sub>15</sub> e (TA)<sub>15</sub> (verde) em metáfases de fêmeas de espécies de Parodontidae: Em (a) *A. piracicabae*, (b) *P. nasus*, (c) *Apareiodon* sp., (d) *P. hilarii* e, (e) *A. affinis*.....116
- Fig. Suplementar 1a** TIRs do elemento transponível En/Spm no genoma macho de *Apareiodon* sp. similares às sequências (TIR-5' CACAGCAAAA.....AGTGT.AA.T....T e TIR3' GTGTCGTTTT.....CACAATT.....). Análise dos 30 pares de bases das extremidades das sequências (destaque em vermelho). Numeração à esquerda refere-se à identificação de cada sequência no genoma. As regiões conservadas (5' CACAGCAAAA.....AGTGT.AA.T....T e 3' GTGTCGTTTT.....CACAATT.....) estão destacadas em azul.....120
- Fig. Suplementar 1b** TIRs do elemento transponível En/Spm no genoma fêmea de *Apareiodon* sp. (TIR-5' CACAGCAAAA.....AGTGT.AA.T....T e TIR3' GTGTCGTTTT.....CACAATT.....). Análise dos 30 pares de bases das extremidades das sequências (destaque em vermelho). Numeração à esquerda refere-se à identificação de cada sequência no genoma. As regiões conservadas (5' CACAGCAAAA.....AGTGT.AA.T....T e 3' GTGTCGTTTT.....CACAATT.....) estão destacadas em azul.....121
- Fig. Suplementar 1c** TIRs do elemento transponível En/Spm no genoma macho de *Apareiodon* sp. similares às sequências (TIR-5' CACTGTAAAA e TIR-3' GTGACATTTT). Análise dos 30 pares de bases das extremidades das sequências (destaque em vermelho). Numeração à esquerda refere-se à identificação de cada sequência. As regiões conservadas (5' CACTGTAAAA e 3' GTGACATTTT) estão destacadas em azul.....122
- Fig. Suplementar 1d** TIRs do elemento transponível En/Spm no genoma fêmea de *Apareiodon* sp. similares às sequências (TIR-5' CACTGTAAAA e TIR-3' GTGACATTTT). Análise dos 30 pares de bases das extremidades das sequências (destaque em vermelho). Numeração à esquerda refere-se à identificação de cada sequência no genoma. As regiões conservadas (5' CACTGTAAAA e 3' GTGACATTTT) estão destacadas em azul.....123
- Fig. suplementar 2** Análise de densidade de repetições *in tandem* em *contigs* contendo mais de 10 sequências En/Spm. Em (a, b) *contigs* 1553664 e 1610658 do genoma macho contendo alta densidade de repetições *in tandem*; em (c – e) *contigs* 1031971, 235765 e 3250488 do genoma fêmea contendo alta densidade de repetições *in tandem*. As linhas diagonais em azul representam fragmentos curtos (< que 100 pb) e as linhas diagonais em vermelho representam unidades > que 100 pb.....124



## LISTA DE TABELAS

<b>TABELA 1.</b> Espécies da família Parodontidae estudadas, localidade, distribuição em bacias hidrográficas e número de espécies.....	<b>50</b>
---	-----------

### CAPÍTULO 1

<b>Table 1</b> Whole and degenerate protein domains of Gypsy detected in <i>Apareiodon</i> sp. Data based on 121 contigs with size $\geq 1900$ bp.....	<b>73</b>
<b>Supplementary Table 1</b> Primer sets constructed for the amplification and characterization of the RTE-LTR sequences Gypsy, using the software Primer3Plus. Expected fragment size $\sim 950$ bp for each set.....	<b>79</b>
<b>Supplementary Table 2</b> Primer sets constructed for the expression analysis of the RTE-LTR sequences Gypsy, using the software Primer3Plus.....	<b>80</b>
<b>Supplementary Table 3</b> Parodontidae species name, sampling locality, distribution in river basins and number of species studied.....	<b>80</b>
<b>Supplementary Table 4</b> Description of the protein domains found in the 145 contigs analyzed from <i>Apareiodon</i> sp.....	<b>81-84</b>

### CAPÍTULO 2

<b>Tabela 1</b> Predição de domínios proteicos detectados nas sequências que apresentaram similaridade ao elemento En/Spm dos genomas macho/fêmea de <i>Apareiodon</i> sp.....	<b>112</b>
<b>Tabela 2</b> Tamanho e número de repetições microssatélites GA/CT e TA/AT detectadas nas sequências En/Spm dos genomas macho e fêmea <i>Apareiodon</i> sp.....	<b>113</b>
<b>Tabela Suplementar 1</b> Nome das espécies de Parodontidae, localidade de amostragem, distribuição em bacias hidrográficas e número de espécies estudadas.....	<b>117</b>
<b>Tabela Suplementar 2a</b> Contigs contendo mais de uma sequência En/Spm em genoma de macho.....	<b>118</b>
<b>Tabela Suplementar 2b</b> Contigs contendo mais de uma sequência En/Spm em genoma de fêmea.....	<b>119</b>

## LISTA DE ABREVIATURAS

**AP:** Proteinase Aspártica  
**CA:** Capsídeo  
**CHR:** Cromo  
**CTAB:** Brometo de cetiltrimetilamónio  
**DBP:** Proteínas de ligação ao DNA  
**DIRs:** Elemento transponível *Dictyostelium*  
**ERVs:** Endógenos retrovírus  
**FISH:** Hibridização *in situ* fluorescente  
**GAG:** Antígeno específico do grupo  
**Gb:** Gigabase  
**Indels:** Inserções e deleções  
**INT:** Integrase  
**Kb:** Kilobase  
**KCl:** Cloreto de potássio  
**LINE:** Elementos nucleares intercalados longos  
**LTR:** Longa região terminal  
**MA:** Matriz  
**m-sm:** Metacêntricos/submetacêntrico  
**MITES:** Elementos transponíveis de repetição invertida em miniatura  
**µl:** Microlitro  
**µM:** Micromolar  
**mM:** Milimolar  
**NC:** Nucleocapsídeo  
**NGS:** Sequenciamento de próxima geração  
**ORF:** Fase aberta de leitura  
**Pb:** Pares de bases  
**PBS:** Sítio de ligação para iniciadores  
**pH:** Potencial hidrogeniônico  
**PLE:** Retrotransposons tipo Penélope  
**PNMA:** Antígenos Paraneoplásicos Ma  
**Pol:** Poliproteína  
**PPT:** Trato de polipurina  
**RPA:** Região pseudoautossômica  
**REW:** Região específica do cromossomo W  
**RH:** Rnase H  
**RON:** Região organizadora de nucléolo  
**RT:** Transcriptase reversa  
**RTE:** Elemento retrotransposon

**SINE:** Elementos nucleares intercalados curtos  
**SNP:** Polimorfismo de nucleotídeo único  
**SSR:** Repetição de sequências simples  
**St:** Subtelocêntricos  
**SSC:** Solução tampão de citrato de sódio  
**STR:** Repetições em tandem curtas  
**TFBS:** Sítio de ligação para fatores de transcrição  
**TE:** Elementos transponíveis  
**TIRs:** Regiões invertidas terminais  
**TRIS:** Hidroximetil-aminometano  
**TSD:** Duplicação do sítio alvo  
**U:** Unidade  
**WGD:** Duplicação do genoma inteiro

## LISTA DE SÍMBOLOS

$\beta$  – beta

$^{\circ}\text{C}$  – Grau Celsius

TM – marca comercial (do inglês *Trade Mark*)

® – marca registrada

$\geq$  – maior ou igual

$\sim$  – valor aproximado

# SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>21</b>
<b>2 REVISÃO DE LITERATURA .....</b>	<b>23</b>
2.1 DNA REPETITIVO .....	23
2.1.1 DNAs satélites, minissatélites e microssatélites .....	23
2.1.2 Elementos repetitivos dispersos .....	26
2.1.2.1 Elementos de Classe I: retrotransposons .....	28
2.1.2.2 Elementos de Classe II: transposons de DNA .....	29
2.2 O IMPACTO E O PAPEL EVOLUTIVO DOS ELEMENTOS TRANSPONÍVEIS NOS GENOMAS .....	31
2.3 ELEMENTOS TRANSPONÍVEIS EM PEIXES TELEÓSTEOS .....	35
2.4 ANÁLISES EM LARGA ESCALA .....	37
2.5 A FAMÍLIA PARODONTIDAE .....	38
2.5.1 Classificação, distribuição e características morfológicas.....	38
2.5.2 Citogenética clássica e molecular.....	41
2.5.2 Estudos citogenéticos e moleculares envolvendo os cromossomos sexuais .....	43
<b>3 OBJETIVOS .....</b>	<b>48</b>
3.1 GERAL.....	48
3.2 ESPECÍFICOS .....	48
<b>4 MATERIAL E MÉTODOS .....</b>	<b>49</b>
4.1 MATERIAL.....	49
4.2 METODOLOGIA .....	50
4.2.1 Caracterização molecular retrotransposon Gypsy e Transposon de DNA En/Spm.....	52
4.2.2 Identificação e análise de DNA satélites .....	52
4.2.3 Construção de primers .....	52
4.2.4 Extração de DNA genômico.....	52
4.2.5 Amplificação das sequências Gypsy e En/Spm para sequenciamento e localização in situ .....	52
4.2.6 Purificação e sequenciamento dos fragmentos Gypsy e En/Spm.....	53
4.2.7 Marcação de sondas .....	53
4.2.8 Preparações mitóticas.....	53
4.2.9 Preparação das lâminas .....	54
4.2.10 Hibridação in situ fluorescente (FISH) .....	54
4.2.11 Análises dos resultados e montagem dos cariótipos.....	54
4.2.12 Extração de RNA e análise de expressão.....	55
<b>5 RESULTADOS.....</b>	<b>56</b>
<b>5.1 CAPÍTULO 1 .....</b>	<b>57</b>



<b>ANCIENT INVASION PHASE, MOLECULAR DEGENERATION AND PROBABLE CO-OPTATED PNMA OF RETROTRANSPOSON GYPSY IN THE GENOME OF A NEOTROPICAL PARODONTIDAE FISH .....</b>	<b>57</b>
<b>1 INTRODUCTION.....</b>	<b>58</b>
<b>2 MATERIALS AND METHODS .....</b>	<b>60</b>
2.1 NEXT GENERATION SEQUENCING (NGS) AND GENOME ASSEMBLY .....	60
2.2 IDENTIFICATION, LANDSCAPE ANALYSIS AND AMPLIFICATION OF GYPSY SEQUENCES .....	60
2.3 MOLECULAR CHARACTERIZATION OF GYPSY .....	61
2.4 RNA EXTRACTION AND TRANSCRIPTION ANALYSIS .....	62
2.5 IN SITU LOCATION .....	62
<b>3 RESULTS .....</b>	<b>63</b>
3.1 GENOME ASSEMBLY, REPETITIVE FRACTION IDENTIFICATION AND REPEAT LANDSCAPE ANALYSIS.....	63
3.2 SIMILARITY ANALYSIS .....	64
3.3 GYPSY SEQUENCES ANALYSIS .....	64
3.4 GENE EXPRESSION ANALYSIS.....	66
3.5 IN SITU LOCATION .....	66
<b>4 DISCUSSION .....</b>	<b>66</b>
<b>REFERENCES.....</b>	<b>72</b>
<b>5.2 CAPÍTULO 2 .....</b>	<b>93</b>
<b>ELEMENTO MÓVEL EN/SPM SENESCENTE EM UM GENOMA DE PEIXE NEOTROPICAL: GÊNESE DE REPETIÇÕES IN TANDEM E MECANISMOS MOLECULARES NA DIFERENCIAÇÃO DOS CROMOSSOMOS SEXUAIS ZW .....</b>	<b>93</b>
<b>1 INTRODUÇÃO.....</b>	<b>94</b>
<b>2 MATERIAL E MÉTODOS .....</b>	<b>96</b>
2.1 SELEÇÃO E CARACTERIZAÇÃO DE SEQUÊNCIAS EN/SPM .....	96
2.2 CONSTRUÇÃO DO REPEAT LANDSCAPE DE EN/SPM .....	97
2.3 CONTAGEM DE SEQUÊNCIAS MICROSSATÉLITES NAS SEQUÊNCIAS EN/SPM .....	97
2.4 AVALIAÇÃO DE REPETIÇÕES IN TANDEM EM CONTIGS CONTENDO SEQUÊNCIAS EN/SPM .....	97
2.5 OBTENÇÃO DE SONDAS PARA A LOCALIZAÇÃO IN SITU.....	98
2.6 PREPARAÇÕES CROMOSSÔMICAS E LOCALIZAÇÃO IN SITU .....	99
<b>3 RESULTADOS.....</b>	<b>100</b>
3.1 IDENTIFICAÇÃO, SELEÇÃO E CARACTERIZAÇÃO DO ELEMENTO EN/SPM .....	100
3.2 ANÁLISE GRÁFICA DA INVASÃO DO GENOMA PELO TRANSPOSON EN/SPM .....	101
3.3 IDENTIFICAÇÃO DE SSR (GA/CT) <sub>N</sub> E (TA/AT) <sub>N</sub> NAS SEQUÊNCIAS EN/SPM DE <i>APAREIODON</i> SP. ....	101
3.4 PRESENÇA DE REPETIÇÕES IN TANDEM EM CONTIGS CONTENDO SEQUÊNCIAS EN/SPM .....	101

3.5 AMPLIFICAÇÃO, SEQUENCIAMENTO E ANÁLISE DE SIMILARIDADE DA SEQUÊNCIA EN/SPM .....	102
3.6 HIBRIDIZAÇÃO IN SITU FLUORESCENTE .....	102
<b>4 DISCUSSÃO .....</b>	<b>103</b>
4.1 DETERIORAÇÃO MOLECULAR DO ELEMENTO EN/SPM NO GENOMA DE <i>APAREIODON SP.</i> .....	104
4.2 CO-OPÇÃO MOLECULAR DO ELEMENTO EN/SPM .....	106
4.3 GÊNESE DE SEQUÊNCIAS MICROSSATÉLITES INTERNAS AO ELEMENTO EN/SPM .....	106
4.4 PROPOSTA DE GÊNESE DE DNA SATÉLITE A PARTIR DO TE EN/SPM .....	108
4.5 DIFERENCIAÇÃO DOS CROMOSSOMOS SEXUAIS .....	109
<b>5 CONCLUSÃO .....</b>	<b>110</b>
<b>REFERÊNCIAS .....</b>	<b>111</b>
<b>6 CONSIDERAÇÕES FINAIS .....</b>	<b>125</b>
<b>REFERÊNCIAS .....</b>	<b>126</b>
<b>ANEXO 1. LICENÇA PERMANENTE PARA COLETA DE MATERIAL ZOOLOGICO MMA/IBAMA/SISBIO. ....</b>	<b>152</b>
<b>ANEXO 2. CARTA DE APROVAÇÃO PELO COMITÊ DE ÉTICA EM EXPERIMENTAÇÃO ANIMAL DA UNIVERSIDADE ESTADUAL DE PONTA GROSSA. ....</b>	<b>153</b>
<b>ANEXO 3. CONJUNTOS DE PRIMERS SINTETIZADOS PARA VALIDAÇÃO RTE GYPSY E UM CONJUNTO PARA O TRANSPOSON EN/SPM. ....</b>	<b>154</b>
<b>ANEXO 4. EXTRAÇÃO DE DNA - PROTOCOLO ADAPTADO DE DOYLE E DOYLE (1990). ....</b>	<b>155</b>
<b>ANEXO 5. PARÂMETROS UTILIZADOS PARA AMPLIFICAÇÃO DOS ELEMENTOS GYPSY E EN/SPM. ....</b>	<b>156</b>
<b>ANEXO 6. PURIFICAÇÃO DO PRODUTO DA PCR PELO KIT DA GE HEALTHCARE - ILLUSTRAGFX PCR DNA .....</b>	<b>157</b>
<b>ANEXO 7. MARCAÇÃO DE SONDAS POR NICK TRANSLATION, KIT DIG-NICK TRANSLATION MIX (ROCHE APPLIED SCIENCE) .....</b>	<b>158</b>
<b>ANEXO 8. SUSPENSÃO DE LEVEDURA (CAVALLINI; BERTOLLO, 1988). ....</b>	<b>159</b>
<b>ANEXO 9. PREPARAÇÕES DIRETAS PARA OBTENÇÃO DE CROMOSSOMOS MITÓTICOS (BERTOLLO; CIOFFI; MOREIRA-FILHO, 2015). ....</b>	<b>160</b>
<b>ANEXO 10. HIBRIDAÇÃO IN SITU FLUORESCENTE (FISH) SEGUINDO PINKEL ET AL. (1986). ....</b>	<b>161</b>
<b>ANEXO 11. CONJUNTOS DE PRIMERS PARA ANÁLISE DE EXPRESSÃO DO ELEMENTO GYPSY EM <i>APAREIODON SP.</i> .....</b>	<b>163</b>
<b>ANEXO 12. PROTOCOLO DE EXTRAÇÃO DE RNA POR TRIZOL (INVITROGEN) .....</b>	<b>164</b>
<b>ANEXO 13. PROTOCOLO DE SÍNTESE DE DNA COMPLEMENTAR (CDNA) - KIT DA GE HEALTHCARE FIRST-STRAND CDNA SYNTESIS KIT .....</b>	<b>165</b>

## 1 INTRODUÇÃO

A descoberta dos elementos transponíveis (TEs) em 1940, por Bárbara McClintock, representou um divisor de águas no campo da genética. Inicialmente foram considerados apenas como “DNA lixo” (OHNO, 1972), sem influência nos genes, apenas dotados da capacidade de invadir genomas devido a sua capacidade de transposição (BIEMONT, 2010).

Os TEs tiveram um reconhecimento evolutivo nos genomas quando foram reconhecidos quanto a sua influência sobre as taxas de recombinação e rearranjos cromossômicos, seu papel como mutadores, reguladores de genes, como formadores de DNAs satélites, e a possibilidade de serem “domesticados pelo genoma hospedeiro”, passando a desempenhar novas funções genômicas (DORER; HENIKOFF, 1994; FANTI et al., 1998; RAMSAY, et al., 1999; KAPITONOV; JURKA, 1999; VOLFF, 2006; SINZELLE; IZSVÁK; IVICS et al., 2009).

Assim, a sua relevância foi progressivamente incorporada pela comunidade científica ao longo das décadas, mas nem as mentes mais férteis poderiam imaginar todas as suas propriedades e seu enorme impacto na composição e funcionamento do genoma, mudando consideravelmente nosso olhar e compreensão das maneiras pelas quais os genomas evoluem e funcionam (RAMOS et al., 2007; FESCHOTTE, 2008; BIEMONT, 2010).

Embora ainda estejamos longe de ter uma visão clara ou completa das interações e regulação da rede gênica que os elementos transponíveis e outras sequências repetidas estão envolvidas, diante da complexidade dos genomas eucarióticos, as novas tecnologias de sequenciamento e análises de genoma têm contribuído de forma significativa para elucidar os mais diversos eventos nos quais os elementos repetitivos possam estar envolvidos (MACAS et al., 2011; VALENTE et al., 2014).

Em vertebrados, os elementos repetitivos têm sido extensivamente explorados, mas, especificamente em peixes neotropicais os estudos são ainda escassos (MEDSTRAND et al., 2005; WARREN et al., 2015). Nos últimos anos, os estudos de mapeamento cromossômico identificaram uma diversidade de elementos repetitivos presentes no genoma das espécies de Parodontidae. Foram identificados acúmulos de DNAs satélites e de TEs no cromossomo sexual W, além de cópias autossômicas destes elementos, atribuindo aos DNAs repetitivos um papel importante

na diversidade e evolução cariotípica observada na família (BELLAFRONTÉ et al. 2011; SCHEMBERGER et al., 2014, 2016; ZIEMNICZAK et al., 2014; TRALDI et al., 2016; NASCIMENTO et al., 2018).

Atualmente, estudos com espécies de Parodontidae vêm buscando estratégias mais resolutivas para uma melhor compreensão dos mecanismos envolvidos na evolução cariotípica, principalmente no que se refere aos elementos repetitivos. Neste sentido, dados de sequenciamento de DNA em larga escala identificaram que 36% do genoma de *Apareiodon* sp. é composto por DNAs repetitivos de diferentes classes, principalmente por DNA transposon e microssatélites (SCHEMBERGER et al., submetido à publicação). No entanto, estudos de caracterização molecular destes elementos ainda são restritos (SCHEMBERGER et al., 2016).

Assim, este estudo apresenta uma análise de diferenciação e distribuição genômica de DNAs repetitivos com foco na evolução cariotípica de *Apareiodon* sp. a partir dos dados de sequenciamento de DNA em larga escala, caracterização molecular e localização *in situ*, contribuindo para uma melhor compreensão de mecanismos específicos de diferenciação cariotípica e genômica em espécies de Parodontidae.

## 2 REVISÃO DE LITERATURA

### 2.1 DNA REPETITIVO

Os organismos eucariontes são complexos quanto à estrutura e organização genômica. Seus cromossomos são compostos por sequências únicas e sequências repetitivas (WAGNER; MAGUIRE; STALLINGS, 1993). As sequências repetitivas são fragmentos de DNA que se repetem diversas vezes e são abundantes nos genomas eucarióticos da maioria das espécies, representando cerca de 80% do genoma de plantas e anfíbios (LEWIN, 2014), 50% ou mais do genoma humano (THE GENOME INTERNATIONAL SEQUENCING CONSORTIUM, 2001) e 36% em um genoma de peixe Neotropical (SCHEMBERGER et al., submetido à publicação).

Os DNAs repetitivos são classificados em sequências moderadamente ou altamente repetitivas de acordo com o seu grau de repetitividade, podendo ainda ser agrupados de acordo com outros critérios, como sua organização (em tandem ou dispersos) ou seu papel funcional. Entre os DNAs repetitivos *in tandem* estão os genes de RNAs ribossômicos e as famílias de gênicas das histonas e globinas, DNA telomérico, bem como os satélites, incluindo DNA centromérico, minissatélites e microssatélites. Já as repetições dispersas agrupam os transposons de DNA e retrotransposons (LÓPEZ-FLORES; GARRIDO-RAMOS, 2012).

#### 2.1.1 DNAs satélites, minissatélites e microssatélites

Os DNAs satélites são sequências altamente repetitivas, geralmente constituídos de 100 a 300 pares de bases (pb) de comprimento (podem ocorrer maiores e/ou menores) e organizados em longas matrizes composta de 1.000 a mais de 100.000 cópias no genoma. Geralmente, localizam-se nas regiões terminais e centroméricas de um ou mais locos cromossômicos, sendo o principal componente da heterocromatina (MARTINS, 2007; PLOHL et al., 2008). Os DNAs satélites podem representar até 50% de alguns genomas de insetos e roedores, 20% do DNA nuclear da planta e menos de 5% em humanos (MACAS et al., 2010; LANDER, et al., 2001).

Nos últimos anos, tem sido sugerida a importância dos DNAs satélites associados com proteínas específicas, nos processos da estabilidade da estrutura genômica, efeitos sobre a regulação gênica, no pareamento cromossômico e



segregação, atuando em função estrutural, assim como na investigação da variabilidade e diversidade genética envolvendo aspectos evolutivos (O'CONNELL; WRIGTH, 1997; KURENOVA et al., 1998; MONOD et al., 2002; HERNÁNDEZ-HERNÁNDEZ et al., 2008; BUSSIEK et al., 2009).

Os DNA satélites podem se originar pelo mecanismo da replicação saltatória (BRITTEN; KOHNE, 1966; 1968), por *crossing-over* desigual (SMITH, 1974) e também a partir de cópias degeneradas de elementos transponíveis (HEIKKINEN et al., 1995). Quanto a sua localização, os DNAs satélites podem estar organizados em sítios específicos de um par cromossômico, podendo formar unidades de repetição que evoluem em paralelo, levando a uma homogeneização intraespecífica entre as unidades repetitivas (DOVER, 1986; UGARKOVIC; PLOHL, 2002). Esta distribuição cromossômica é reflexo da evolução em concerto de sequências entre cromossomos não homólogos onde a dispersão intercromossômica seria mais rápida que as taxas de mutação e/ou homogeneização intracromossômica (PLOHL et al., 1992).

Diante da dinâmica evolutiva extremamente variável dos DNAs satélites, estes podem, de fato, prover informações úteis tanto para estudos espécie específicos ou cromossomo específicos (UGARKOVIC; PLOHL, 2002), como para estudos filogenéticos (PONS; JUAN; PETITPIERRE, 2002; PONS; GILLESPIE, 2003; KANTEK et al., 2009). No entanto, famílias de DNA satélite podem, paralelamente, apresentar evolução lenta e permanecer conservadas ao longo do tempo com reflexo na homogeneização das sequências (DE LA HERRÁN et al., 2001).

Os minissatélites são sequências geralmente constituídas de 1 a 5 kilobase (Kb) e consistem em números variáveis de unidade repetida (usualmente de 15 a 100 nucleotídeos de comprimento). O DNA de minissatélite já foi localizado nas regiões subteloméricas dos cromossomos de humanos, em *Caenorhabditis elegans* e em *Tetraodon nigroviridis*, enquanto que em *Arabidopsis thaliana* estas repetições tendem a se agrupar na região pericentromérica e, em *Saccharomyces cerevisiae* na região centromérica e subtelomérica (ROEST-CROLLIUS et al., 2000; VERGNAUD; DENOEUDE, 2000; RICHARD; DUJON, 2006).

Os microssatélites, também são conhecidos como repetições de sequência simples (SSRs) ou repetições curtas em tandem (STRs) (WEBER; MAY, 1989). Estes são geralmente constituídos de unidades de 1 a 6 nucleotídeos que se repetem de 5 a 40 vezes em um determinado loco e, em alguns casos podem apresentar número

superior de repetições (WAN et al., 2004; LITT; LUTY, 1989; TAUTZ, 1989; WEBER; MAY, 1989).

Os microssatélites podem ser encontrados tanto em regiões gênicas codificadoras quanto em regiões não codificantes, incluindo as sequências reguladoras. Dinucleotídeos são os tipos mais representados de repetições microssatélites para muitas espécies, embora repetições contendo múltiplos de 3 nucleotídeos são as mais abundantes em regiões de codificação, presumivelmente porque elas não causam uma mutação de mudança no quadro de leitura (LÓPEZ-FLORES; GARRIDO-RAMOS, 2012).

Quanto a sua densidade, nos genomas humano e de camundongo, a localização dos microssatélites é quase duas vezes maior perto das extremidades dos braços cromossômicos (BACHTROG et al., 1999). Em várias espécies, a densidade e/ou a distribuição de microssatélites no cromossomo X difere daquela nos autossomos, podendo ser associado a fatores como diferenças sexuais na taxa mutação, diferenças no tamanho da população efetiva entre o cromossomo X e autossomos, e a eficiência de seleção em cromossomos hemizigóticos (BACHTROG et al., 1999).

Estima-se que a proporção de sequências de microssatélites em plantas seja inferior a 1% do tamanho genômico, enquanto para as espécies de peixes *Fugu rubripes*, *Tetraodon nigroviridis* e *Apareiodon* sp. as estimativas são de 2,12%, 3,21% e 3,4% respectivamente (ROEST-CROLLIUS et al., 2000; MORGANTE; HANAFEY; POWELL, 2002; SCHEMBERGER et al., submetido à publicação).

A evolução de um microssatélite pode ser representada por um “ciclo de vida” (AMÓS 1999; BUSCHIAZZO; GEMMELL 2006), composto por: (i) *nascimento*, quando um lóco adquire o número mínimo de repetições necessárias para altas taxas de derrapagem (*slipage*) da maquinaria de replicação; (ii) *idade adulta*, caracterizada por aumento no número de repetições e (c) *morte*, quando um lóco degenera para um número de repetição abaixo do limite, deixando de sustentar altas taxas de derrapagem durante a duplicação da cromatina (BUSCHIAZZO; GEMMELL, 2006).

Os elementos transponíveis (TEs) também podem atuar no ciclo de vida de microssatélites. Estudo realizado com primatas mostra que 36,3% dos microssatélites presentes em posições ortólogas em humanos, chimpanzé e orangotango provavelmente foram adquiridos devido à inserção TEs (KELKAR et al., 2011). Entre os nascimentos e mortes de microssatélites que ocorreram entre uma linhagem

ancestral comum de humano e chimpanzé, 26,3% dos nascimentos dos microssatélites e, 24.3% das suas mortes ocorreram dentro de TEs após sua integração (KELKAR et al., 2011).

Neste sentido é sugerida uma relação mutualística entre microssatélites e TEs em favor de uma coevolução, onde as repetições em tandem são postuladas para facilitar a transposição, servindo como uma ferramenta, direcionando a interação homóloga em sequências-alvo para integração (HEIKKINEN et al., 1995; NADIR, et al., 1996). Em contrapartida, a degeneração molecular de elementos transponíveis contribuem para a origem de sequências satélites (HEIKKINEN et al., 1995; NADIR, et al., 1996).

### 2.1.2 Elementos repetitivos dispersos

Elementos repetitivos dispersos, também conhecidos como elementos genéticos móveis ou elementos transponíveis são segmentos de DNA que possuem como particularidade a capacidade de se mover no genoma. Eles possuem também características de “parasitas moleculares” por se aproveitarem da estrutura proporcionada pela célula, promovendo a sua própria duplicação, independente do ciclo celular (CAPY, et al., 1998; KIDWELL; LISCH, 2001; CRAIG, 2002).

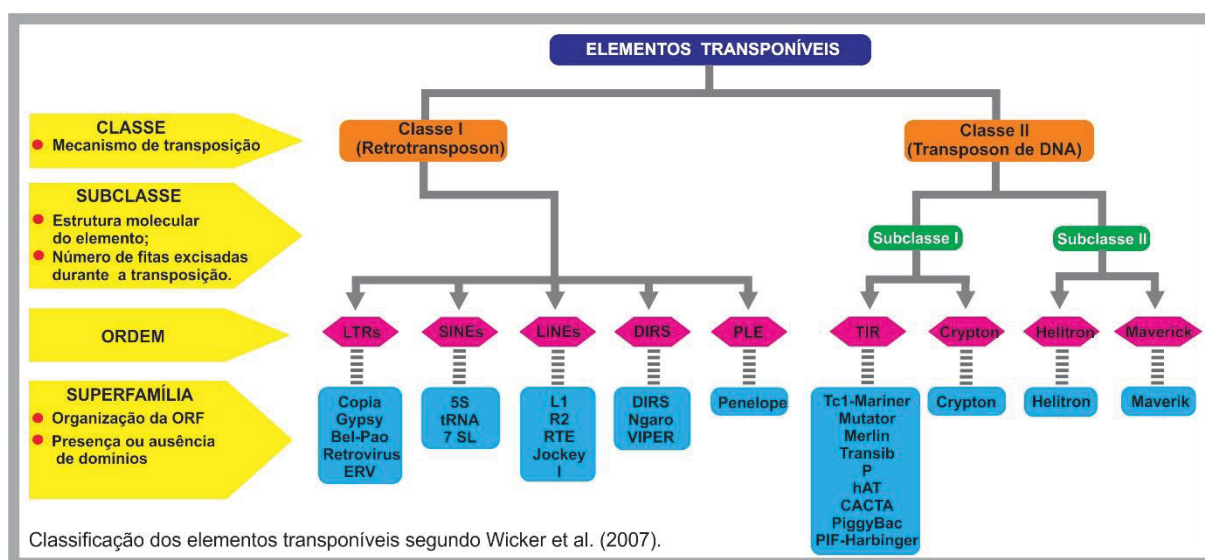
Os TEs foram identificados na década de 50, por Bárbara McClintock denominando-os como “elementos controladores” sugerindo que tais elementos desempenhavam um papel relevante na regulação gênica e, sem desconsiderar os possíveis efeitos danosos da ocorrência da mobilização destes elementos do genoma (MCCLINTOCK, 1950; KIDWELL; LISCH, 2001).

Devido a grande quantidade, diversidade e mecanismos de transposição os TEs foram classificados de acordo com tipos de transposição e similaridade das sequências. No entanto, este assunto ainda está em debate, não possuindo uma classificação universal. Neste estudo adotamos o sistema de Wicker et al. (2007), onde estes elementos repetitivos são classificados em duas classes principais de TEs, baseado no seu mecanismo de transposição:

- *Classe I ou retrotransposon*: Elementos que se movem por transcrição reversa, através de um intermediário de RNA;
- *Classe II ou DNA transposons*: Elementos com capacidade de excisão do seu local original e inserção em outro lugar no genoma.

Os TEs ainda são divididos em subclasses de acordo com a estrutura molecular do elemento ou pelo número de fitas excisadas durante a transposição. Em um terceiro nível hierárquico são agrupados em superfamílias baseado na organização das suas fases aberta de leitura (ORFs - *open reading frame*), presença ou ausência de domínios e assinatura proteica (Figura 1). Por último, são organizados em famílias e subfamílias de acordo com características mais específicas dos elementos, principalmente no que se refere a integridade das sequências de DNA (WICKER et al., 2007).

FIGURA 1. Classificação dos elementos transponíveis.



FONTE: O autor (2019).

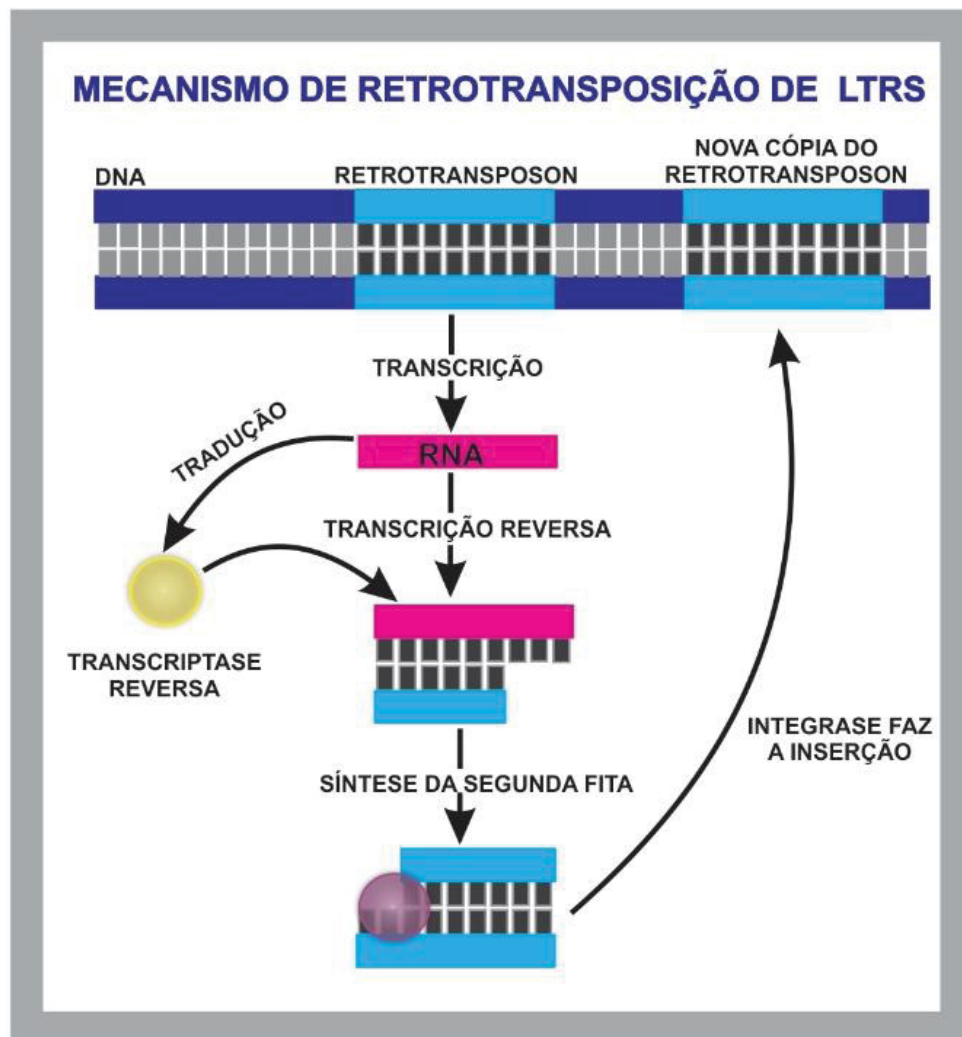
A transposição de um TE, independente da sua classe, pode ocorrer de forma autônoma quando são capazes de sintetizar as enzimas necessárias para sua transposição e, não autônomos, quando não possuem a capacidade de codificação, podendo ser subsidiados pelas enzimas dos autônomos durante a sua mobilidade (CAPY, et al., 1998; HUA-VAN et al., 2011).

A maioria dos TEs, exceto Crypton, Helitron e DIRS apresentam, flanqueando as suas extremidades, uma duplicação do sítio alvo (TSD). Estas sequências de 2-11pb, diretamente repetidas, são produzidas devido aos dois filamentos do DNA alvo serem clivados em posições diferentes (corte coesivo) e, consequente inserção do TE entre as extremidades unifilamentares. Assim, a lacuna unifilamentar formada será preenchida pelo mecanismo de reparo, gerando uma duplicação direta do sítio alvo (WICKER et al., 2007; SNUSTAD; SIMMONS, 2013).

### 2.1.2.1 Elementos de Classe I: retrotransposons

Os elementos retrotransposons (*RTE-retrotransposable elements*) se movimentam no genoma por meio de um intermediário de RNA que é transcrito reversamente para DNA pela ação da enzima transcriptase reversa e integrado em uma nova localização dentro do genoma hospedeiro. Neste mecanismo os RTEs são copiados de seu local de origem, mantendo o sítio doador intacto e, por isso, frequentemente seu mecanismo é referido como uma transposição do tipo "copiar e colar" (FIGURA 2, EICKBUSH; JAMBURUTHUGODA, 2008).

FIGURA 2. Representação esquemática do mecanismo de retrotransposição de um elemento LTR.



FONTE: O Autor (2019).

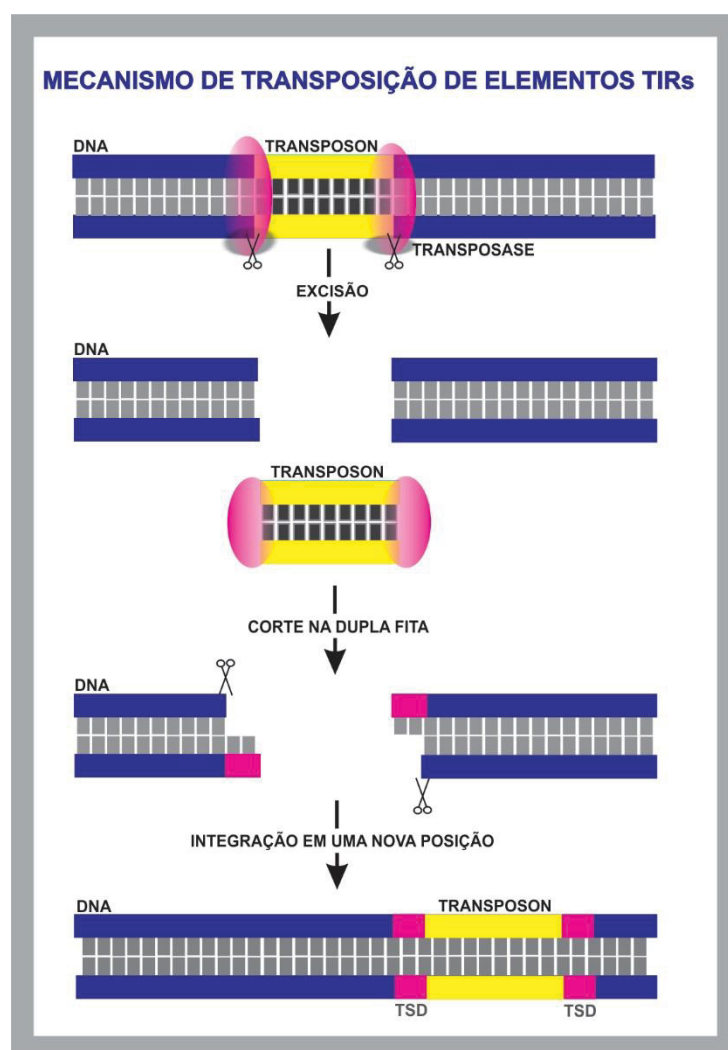
A enzima transcriptase reversa é codificada pelo retrotransposon e a sua estrutura molecular e filogenética é utilizada para alocação dos elementos nas cinco ordens, a saber: LTR (longas repetições terminais), LINE (elementos nucleares intercalados longos), SINE (elementos nucleares intercalados curtos), DIRS (sequência de repetição intermediária de *Dictyostelium*) e PLE (elementos semelhantes a Penelope) (WICKER et al., 2007).

Quanto ao mecanismo de retrotransposição em RTE-LTRs, a molécula de DNA originada pela transcriptase reversa fica livre no meio celular e segue para a inserção em um novo local pela ação da enzima integrase (HAVECKER; GAO; VOYTAS, 2004). Nos elementos não-LTRs, o RNA intermediário ancora-se ao DNA-alvo antes de ser retrotranscrito em DNA e, neste caso, não há moléculas de DNA livres. Nestes elementos, a integração no genoma ocorre através da enzima endonuclease em elementos LINE e SINE e da tirosina recombinase nos elementos DIRS (OSTERTAG; KAZAZIAN, 2001).

#### 2.1.2.2 Elementos de Classe II: transposons de DNA

Os transposons de DNA ou DNA transposons se movimentam no genoma por meio da clivagem do seu sítio original e inserção em um novo sítio do DNA, sendo estes conhecidos por transposons do tipo “corta e cola” ou, por meio de transposição e cópia da fita de DNA para nova região genômica “copia e cola” (WICKER et al., 2007). Devido a estes mecanismos são subdivididos em duas subclasses de acordo com o número de fitas de DNA que são cortadas durante a transposição (FIGURA 3, WICKER et al., 2007).

FIGURA 3. Representação esquemática do mecanismo de transposição de transposons clássicos



FONTE: O Autor (2019).

Nos elementos da subclasse I, flanqueando a região codificadora são encontradas as repetições invertidas terminais (TIRs) e, para sua transposição ocorre a excisão dos dois filamentos de DNA. Na subclasse II estão presentes os TEs do tipo Helitron, que se replicam pelo mecanismo círculo rolante com excisão de apenas uma fita de DNA e Maverik que se movem provavelmente por um mecanismo via corte de um filamento de DNA seguido de uma replicação extracromossômica e integração a um novo local do genoma (HURST; WERREN, 2001).

Os elementos da subclasse I, que são os transposons clássicos, apresentam pelo menos uma ORF que codifica para a enzima transposase, que medeia a transposição pelo reconhecimento das TIRs e excisão das fitas, mas algumas superfamílias podem apresentar ORFs adicionais codificando para outras enzimas, em sua grande maioria com função desconhecida (WICKER et al., 2007).



O mecanismo de transposição pode ser replicativo quando há a duplicação de uma cópia ou conservativo, que consiste somente em excisão e inserção em um novo local (HURST; WERREN, 2001; WICKER et al., 2007). No mecanismo replicativo, os transposons de DNA, podem aumentar em números de cópias durante duplicação do DNA, se transpondo de uma posição já duplicada para outra onde forquilha de duplicação ainda não tenha passado (NASSIF et al., 1994; CAPY et al., 1998).

## 2.2 O IMPACTO E O PAPEL EVOLUTIVO DOS ELEMENTOS TRANSPONÍVEIS NOS GENOMAS

Há mais de 70 anos, com o trabalho de Barbara McClintock, surgiu a primeira ideia de que os TEs têm um papel fundamental na evolução e controle de genes eucarióticos. Deste então, inúmeros trabalhos buscam elucidar o potencial dos TEs para fornecer aos genomas hospedeiros uma importante fonte de diversidade genética e evolução (BIÉMONT, 2010; CHUONG; ELDE; FESCHOTTE, 2016).

Os elementos móveis têm “preferências peculiares” para sua inserção, visando um ambiente favorável para sua expressão e mobilidade (BIÉMONT et al., 1997). Segundo Biémont et al. (1997), alguns TEs evoluíram mecanismos moleculares sofisticados para os "redutos seguros" no genoma, invadindo regiões distantes de sequências gênicas do hospedeiro, como regiões intergênicas, heterocromáticas ou regiões internas de outros elementos. Em contraste, existem TEs que parecem apresentar certa preferência por sítios de inserção localizados próximos ou dentro das sequências gênicas, em regiões intergênicas ou nos íntrons (BIÉMONT et al., 1997; KIDWELL; LISCH, 1997).

Assim, os impactos no genoma hospedeiro causados pela presença e movimentação dos TEs apresentam uma grande divergência, podendo estar relacionados com o aumento da variabilidade genética, que pode colaborar com a geração de uma melhoria na adaptabilidade dos indivíduos portadores, ou pode gerar danos ao genoma hospedeiro que, por sua vez, passa a controlar a mobilização dos elementos de transposição, chegando a inativá-los (KIDWELL; LISCH, 1997; CAPY et al., 1998). Neste sentido, mecanismos como metilação, mutações e interferência por RNA podem ter surgido como resposta para reduzir a mobilização destes elementos,



impedindo o aumento do número de cópias dos TEs no genoma hospedeiro (HURST; WERREN, 2001; CASTRO; CARARETO, 2004).

O mecanismo de duplicação do RTE-LTR é um exemplo onde a cada inserção são geradas duas cópias idênticas da LTR no genoma do hospedeiro (MAGER; STOYE, 2015). Contudo, após a inserção, esses elementos geralmente sofrem recombinação entre as suas LTRs, com a remoção das regiões de codificação desses elementos, deixando apenas as LTR solitárias e intactas, com sequências reguladoras originais. Estes eventos são bem documentados em humanos, onde os Endógenos Retrovírus (ERVs) ocupam ~ 8% do humano genoma, mas 90% são LTRs solitárias (LANDER et al., 2001). Em contraste, os elementos LINEs geralmente perdem a sua sequência promotora e mantém a região de codificação. Os LINEs representam ~20% genoma humano, mas a grande maioria sofreu a remoção das suas sequências promotoras (LANDER et al., 2001).

Da mesma forma, transposons de DNA são propagados principalmente como repetição em miniatura invertida de elemento transponível (MITes), derivados da eliminação da região interna de elementos autônomos. Previsivelmente, muitos MITes não possuem as sequências promotoras de seu elemento parental (FESCHOTTE; ZHANG; WESSLER, 2002).

No entanto, mesmo que a maioria dos TEs presentes nos genomas se apresente como cópias degeneradas, silenciadas e sem a capacidade de transposição, os TEs são considerados uma das principais fontes de mutações (KIDWELL; LISH, 1997, 2001). Estes elementos podem promover deleções, substituições e inserções de nucleotídeos, gerar rearranjos gênicos e cromossômicos, atuar na inativação gênica e induzir à recombinação desigual (KIDWELL; LISH, 1997, 2001). Ainda, os TEs estão relacionados com a hipermutabilidade gênica, podendo resultar em fenótipos nulos ou letais, ou até alterar padrão de expressão fenotípica quando se inserem em regiões reguladoras de genes (KIDWELL; LISH, 1997, 2001; HERÉDIA, 2002).

Em outra via é postulado que os TEs possam apresentar transmissão horizontal, ou seja, a troca de sequências e/ou genes entre espécies reprodutivamente isoladas por meio de um vetor intermediário (CAPY et al., 1998; VIEIRA et al., 2002; HERÉDIA; LORETO; VALENTE, 2004).

Estudos em *Drosophila* evidenciaram o papel dos TEs na possível origem de polimorfismos para inversões cromossômicas em populações naturais (KRIMBAS;

POWELL, 1992). Este mecanismo pode promover a ruptura de um gene, modificar o padrão de segregação dos genes presentes na inversão, ou ainda, pelo efeito de posição, alterar a expressão de genes vizinhos (CÁCERES et al., 1999). Nesta via pode ocorrer a modificação de sequências regulatórias de genes e consequentemente a alteração da sua expressão, ou ainda, provocar o deslocamento de genes para lugares distantes onde sua expressão será alterada ou silenciada pela proximidade da heterocromatina (revisão em KRIMBAS; POWELL, 1992; CÁCERES et al., 1999; ROHDE, 2000; CÁCERES, PUIG; RUIZ, 2001; KIDWELL; HOLYOAKE, 2001; CASALS, CÁCERES; RUIZ, 2003; PUIG; CÁCERES; RUIZ, 2004).

Efeitos fenotípicos da mobilização dos TEs *P*, *I* e *Hobo* foram observados em divisões celulares da linhagem germinativa de moscas consideradas híbridas (prole do cruzamento entre machos com TE e fêmeas que não o possuem) com a geração de prole estéril, efeito conhecido por disgenesia do híbrido (KIDWELL; KIDWELL; SVED, 1977). Nesta via, não existe um controle da mobilização do elemento nos zigotos, gerando moscas que podem ser total ou parcialmente estéreis devido a atrofia gonadal e, desta forma, pode gerar o isolamento reprodutivo (KIDWELL; KIDWELL, SVED, 1977; HURST; SCHILTHUIZEN, 1998; RIO, 2002).

Em adição, os TEs estão associados a modificações no tamanho do genoma hospedeiro, contribuindo para o fator conhecido como o “Paradoxo do valor C” em uma variedade de organismos. Estudos recentes realizados por GAO et al. (2016) com quatro espécies de teleósteos (Zebrafish, Medaka, Stickleback e *Tetraodon*) evidenciaram que as diferenças no acúmulo de TE, principalmente os DNA transposons, explicam as principais variações de tamanho dos genomas dos teleostei investigados. Ainda, estes autores propuseram que o acúmulo e a movimentação dos TEs desempenharam um papel fundamental na evolução do genoma de teleósteos (GAO et al., 2016).

Em outra perspectiva, os TEs podem ainda ser “domesticados pelo genoma hospedeiro”, passando pelo processo de “exaptação” ou “co-opção” molecular (MILLER; MCDONALD; PINSKER, 1997). É descrito que o “ciclo de vida” dos TEs consiste em: (i) no seu surgimento que pode ser resultante de eventos de mutação ou recombinação de sequências presentes no genoma ou, por meio de transferência horizontal, (ii) uma rápida fase de propagação com o aumento do número de cópias e, (iii) uma fase longa de senescência com sua inativação por processos de degeneração de TEs pelo acúmulo de mutações ou seu silenciamento causado por

mecanismos epigenéticos do genoma do hospedeiro (KIDWELL; LISCH, 1997). Nesta última etapa já não existem elementos autônomos nem aumento do número de cópias, e as sequências não autônomas também são perdidas devido a degeneração das suas sequências (KIDWELL; LISCH, 1997). Então estes podem ser eliminados, permanecerem no genoma como cópias neutras ou, serem co-optados para novas funções no genoma (KIDWELL; LISCH, 1997; FERNÁNDEZ-MEDINA et al., 2012).

Nesta perspectiva, diversos estudos foram realizados, a exemplo os elementos *TART* e *HeT* em *Drosophila* que auxiliam em funções básicas do genoma, como a organização da região terminal dos cromossomos (KIDWELL; LISCH, 1997). Esses elementos se inserem nas regiões teloméricas e realizam uma função semelhante à da enzima telomerase (KIDWELL; LISCH, 1997). Outro exemplo a ser citado é da família SINE (SINE- LF) com atividade na regulação gênica identificada em virtude da conservação evolutiva em tetrápodes (BEJERANO et al., 2006). Em ensaios utilizando molécula repórter em camundongos, cópias SINE- LF mostraram possuir atividade de elemento regulatório à distância (enhancer) tecido específico (BEJERANO et al., 2006).

A expressão da enzima amilase para compor a saliva é um exemplo clássico de co-opção molecular de TEs em mamíferos a partir da inserção de um LTR retroviral em um ancestral comum de primatas antropoides (TING et al., 1992). Outro caso interessante é atribuído ao endógeno retrovírus humano (HERV-W e HERV-FRD) na expressão endometrial do hormônio prolactina, essencial durante a gravidez em mamíferos (NOORALI et al., 2009; EMERA; WAGNER, 2012; GERLO et al., 2006).

Alguns estudos recentes sugerem que os TEs podem remodelar extensivamente redes gênicas que estão envolvidas em processos específicos, incluindo a compensação de dose (ELLISON; BACHTROG, 2013), imunidade (CHUONG; ELDE; FESCHOTTE, 2016) e desenvolvimento embrionário (GERDES et al., 2016).

Tem sido descrito que eventos epigenéticos como a metilação do DNA, modificações das histonas e, RNA de Interferência, são todos mecanismos interdependentes associados à conformação da cromatina, que podem interromper a atividade ou inatividade dos TEs. Esses processos devem, portanto, desempenhar um papel não só na defesa do genoma contra a invasão por TEs, mas também nas interações complexas envolvidas na regulação gênica ao longo do desenvolvimento (SOBIESKI; EDEN, 1981; HUDA; MARINÓ-RAMIÍREZ; JORDAN, 2010).

Estas observações até agora relatadas para uma gama diversificada de organismos rejuvenesceram algumas das ideias originais propostas por McClintock e reposicionam os elementos de transposição como um potente impulsionador da evolução nas redes gênicas transcricionais em eucariotos.

## 2.3 ELEMENTOS TRANSPONÍVEIS EM PEIXES TELEÓSTEOS

Os TES acompanharam a evolução do genoma eucariótico por milhões de anos, sugerindo um envolvimento na mudança e modelagem genômica. Particularmente em teleósteos, um número crescente de famílias de elementos transponíveis tem sido identificado nos últimos anos, algumas das quais ainda estão ativas (KOGA et al., 2006; SLOTKIN; MARTIENSSEN, 2007; PEREIRA et al., 2015). Estes estudos estão contribuindo para uma melhor entendimento do papel dos TEs nos rearranjos genômicos (KOGA et al., 2006; PEREIRA et al., 2015) e no controle da atividade gênica (SLOTKIN; MARTIENSSEN, 2007).

Os genomas de teleósteos contêm um número superior de tipos de superfamílias de TEs quando comparado os genomas de tetrápodes (GAO et al., 2016). Especificamente os transposons de DNA parecem ser mais abundantes neste grupo, se apresentam espalhados por todo o genoma, enquanto as famílias de retrotransposons são menos frequentes e podem estar agrupadas em áreas específicas dos cromossomos (JIANG et al., 2011; CHALOPIN, et al., 2015; COAN; MARTINS, 2018; YUAN et al., 2018; SCHEMBERGER et al., submetido à publicação).

Inúmeros elementos de diferentes classes já foram identificados em teleósteos, sendo o Tc1-Mariner, Rex-1, Rex-3 e Rex-6 os mais estudados neste grupo (IVICS et al., 1997; GUO et al., 2015; SENE et al., 2015; GAO et al., 2016; SCHEMBERGER et al., 2016; ZHANG et al., 2016; RIBEIRO et al., 2017; PRIMO et al., 2018; PUCCI et al., 2018). Estudos envolvendo elementos LINES-L2, Helitron, hAT, Zisupton, SINE, PiggyBac e Gypsy também já foram relatados (BOER et al., 2007; OLIVEIRA et al., 2003; COCCA; IORIO; CAPRIGLIONE, 2011; ZHOU, 2012; BARBOSA et al., 2015; CAO et al., 2016; GLUGOSKI et al., 2018).

Quanto ao papel dos TEs nos genomas de teleósteos, embora grande parte destes se mostraram presentes como cópias degeneradas nos genomas e sem a capacidade de transposição, é crescente o número de estudos propondo para os TEs um papel de impulsionadores para a diversidade e evolução cariotípica

(SCHEMBERGER et al., 2016; GLUGOSKI et al., 2018; PRIMO et al., 2018). Os TEs podem estar envolvidos em rearranjos cromossômicos, na evolução dos cromossomos sexuais heteromórficos, gerar sítios para crossing-over desigual e para recombinação ectópica, estabelecer ponto suscetíveis à quebra do DNA e fusão Robertsoniana, entre outros (CHALOPIN et al., 2015; GAO et al., 2016; SCHEMBERGER et al., 2016; GLUGOSKI et al., 2018).

Um estudo com espécies da família Corydoradinae mostrou a associação entre o aumento do número de cópias do TE *TC1-IS630-Pogo* com o evento da duplicação do genoma total (WGD – *Whole Genome Duplication*) em espécies de *Corydoras* (MARBURGER et al., 2018). No salmão os elementos repetitivos da superfamília TC1-Mariner foram associados à especiação no grupo em espécies que passaram por WGD (BOER et al., 2007). Além disso, para o Salmão do Atlântico tem sido sugerido que explosões de transposição de elementos repetitivos impulsionaram rearranjos genômicos que auxiliaram o processo de rediploidização genômica (LIEN et al., 2016).

Alguns estudos propuseram a possibilidade de alguns elementos estarem ativos nos genomas de teleósteos. Pereira et al. (2015) evidenciaram baixas taxas de substituição nucleotídica, ORFs intactas e atividade recente do retroelemento Rex-3 em espécies híbridas de Cyprinidae. Em tambacu foi sugerido uma possível participação dos TEs no mecanismo de disgenesia híbrida (BLAUTH et al., 2009) com propagação de elementos Rex ativos relacionados à processos de regulação gênica (PEREIRA et al., 2015). Ainda, CAO et al. (2016) sugeriram a atividade do RTE Gypsy no genoma de *Carpa Jian* a partir da caracterização molecular do elemento, que apresentou integridade das ORFs e expressão da enzima transcriptase reversa nos tecidos do fígado, rim, sangue, músculo e gônadas.

Vale destacar que os estudos referentes aos TEs em teleósteos são em sua maioria relacionados a sua localização *in situ*. Abordagens mais amplas, que envolvem caracterização molecular e entendimento da função genômica dos TEs têm sido realizadas em poucas espécies que já possuem o genoma sequenciado (COAN; MARTINS, 2018; GAO et al., 2016; SCHEMBERGER et al., submetido à publicação).

Neste sentido, estudos de caracterização molecular de TEs em peixes neotropicais estão apenas iniciando com sequenciamento do genoma de *Apareiodon* sp. (SCHEMBERGER et al., submetido à publicação). A análise da fração repetitiva evidenciou que ~36% do genoma de *Apareiodon* sp. é composto por elementos

repetitivos de diferentes classes (SCHEMBERGER et al., submetido à publicação). Um total de 56 famílias de TEs foram identificadas, sendo os transposons de DNA os mais representativos com ~38% da fração repetitiva, seguido dos retrotransposons com ~26,5%. As sequências microssatélites representaram ~9,4% da fração repetitiva e um total de 20 tipos de microssatélites foram detectados (SCHEMBERGER et al., submetido à publicação).

Análises adicionais evidenciaram que as sequências de TEs em *Apareiodon* sp. se encontram em diferentes níveis de degeneração molecular e o surgimento e a senescência de cada superfamília de TEs apresentam uma história particular. Ainda, a análise concentrada em sequências repetitivas acumuladas no cromossomo W da espécie revelou que os transposons CMC Enspm, Hat, Helitron, Tc1-Mariner e Gypsy e os microssatélites (GACA)<sub>n</sub>, (GATA)<sub>n</sub>, (CAA)<sub>n</sub>, (CAT)<sub>n</sub>, (AC)<sub>n</sub>, (CAG)<sub>n</sub>, (CA)<sub>n</sub> e (GA)<sub>n</sub> são os elementos repetitivos mais acumulados em fêmea (SCHEMBERGER et al., submetido à publicação).

## 2.4 ANÁLISES EM LARGA ESCALA

A última década foi marcada por um tremendo progresso em nossa compreensão de como TEs moldam a evolução do genoma. Este grande avanço foi impulsionado principalmente com a chegada de tecnologias sequenciamento em larga escala que incorporaram inovações revolucionárias para enfrentar as complexidades dos genomas (GOODWIN; MCPHERSON; MCCOMBIE, 2016).

Essas tecnologias realizam o sequenciamento de DNA em plataformas capazes de gerar informações de milhões de pares de bases em um único experimento, com considerável economia de tempo e custo por base sequenciada (MARDIS, 2008). Desta forma, a tecnologia de sequenciamento de próxima geração (NGS - *next generation sequence*) tem impulsionado e modificado uma grande diversidade de métodos de investigação biológica que têm utilizado o sequenciamento de DNA/RNA como base. Portanto, plataformas de NGS estão contribuindo para novas áreas de pesquisa, incluindo a investigação de genomas antigos e a caracterização da diversidade biológica (MARDIS, 2008).

Em adição, as tecnologias NGS oportunizaram estudos mais complexos de organismos poliploides, organismos com altas taxas de sequências repetitivas,



estudos genômicos comparativos, identificação de polimorfismos e de genes diferencialmente expressos (SOUSA; CARNEIRO; CARNEIRO, 2009).

Em contrapartida, ao longo dos últimos anos, alguns desafios são encontrados com a tecnologia de sequenciamento DNA em larga escala, principalmente quando se refere à fração repetitiva de genomas de plantas e animais. Grande parte das ferramentas de bioinformática disponível requer a montagem do genoma completo para a identificação da unidade de repetição ou, se baseiam em pesquisas de similaridade utilizando bases de dados de elementos repetitivos conhecidos. No entanto, o número genomas montados ainda é reduzido, e muitos apresentam baixa qualidade nas regiões repetitivas (NOVAK et al., 2013).

Diante deste quadro, a identificação precisa e anotação da fração de TEs em genomas são tarefas desafiadoras devido a sua significativa diversidade, necessitando de inúmeras ferramentas com diferentes abordagens computacionais. Atualmente três estratégias são mais utilizadas para a anotação de sequências repetitivas: baseado em homologia, baseado em assinatura e abordagem *de novo* (LERAT, 2010; BERGMAN, 2007).

As ferramentas baseadas em assinaturas dependem da estrutura típica de um determinado elemento e podem detectar TEs completos, beneficiando a investigação da estrutura, variação e evolução (BARGHINI et al., 2014; WICKER et al., 2018). A abordagem *de novo* permite a classificação parcial de sequências sem conhecimento prévio, o que pode ser especificamente útil na caracterização de novos elementos repetitivos do genoma (BAO; EDDY, 2002; NOVAK; NEUMANN; MACAS, 2010). Para uma anotação abrangente e confiável, adotar abordagens combinadas tem mostrado ser a melhor estratégia na maioria dos estudos (PERMAL; FLUTRE; QUESNEVILLE, 2012; PLATT; BLANCO-BERDUGO; RAY, 2016).

## 2.5 A FAMÍLIA PARODONTIDAE

### 2.5.1 Classificação, distribuição e características morfológicas

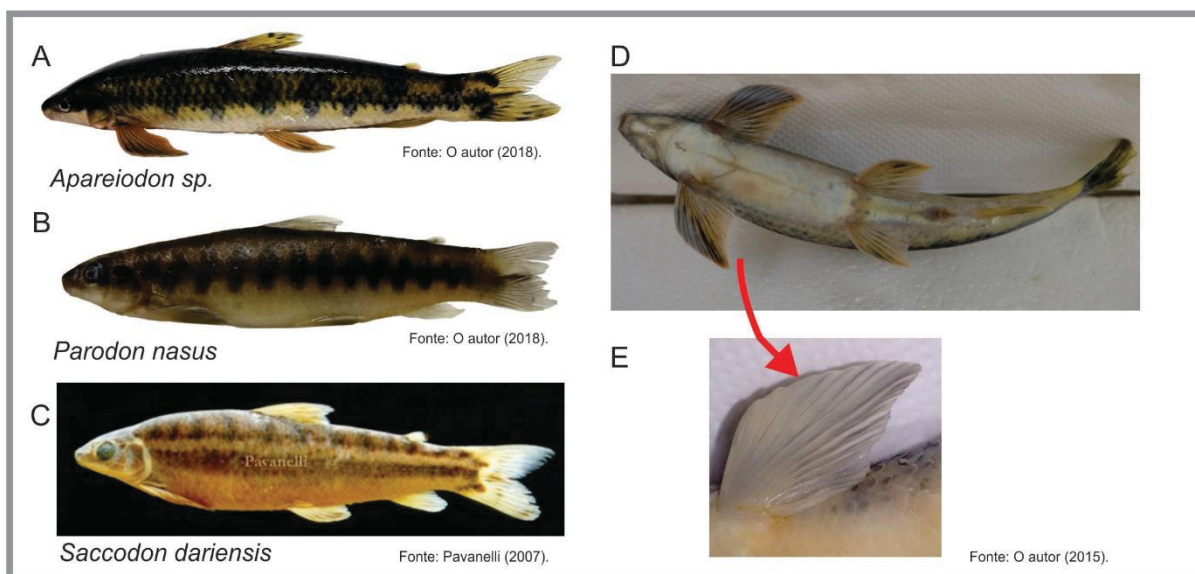
Os peixes de águas doces do Brasil são pertencentes a três linhagens: Chondrichthyes, Actinopterygii e Dipnoi. O grupo Actinopterygii é o mais numeroso e a sua diversidade é distribuída de forma heterogênea, já que apenas cinco ordens detêm mais de 95% das espécies, sendo a ordem Characiformes a mais numerosa,

seguida dos Siluriformes, Perciformes, Cyprinodontiformes e Gymnotiformes. (ROSA; LIMA, 2008).

A ordem Characiformes compreende 18 famílias e, dentre estas, a família Parodontidae consiste em um grupo relativamente pequeno de peixes neotropicais, com 32 espécies válidas (ESCHMEYER; FONG, 2019), divididas em três gêneros: *Parodon* Valenciennes, 1850; *Saccodon* Kner, 1863; e *Apareiodon* Eigenmann, 1916 (PAVANELLI; BRITSKI, 2003). Os gêneros *Parodon* e *Apareiodon* têm distribuição ampla pelo continente sul americano, enquanto *Saccodon* está restrito ao norte do continente, sendo encontrado na bacia costeira no Panamá, Colômbia e Equador (PAVANELLI; BRITSKI, 2003).

A morfologia dos parodontídeos pode ser descrita por um corpo fusiforme cinza pálido (NOMURA, 1979) e coloração composta basicamente de listras, barras e manchas (FIGURA 3A, B e C, SAZIMA, 1980). Nestes peixes, as nadadeiras peitorais e pélvicas são bem desenvolvidas e adaptadas para a estabilização do peixe sobre os fundos rochosos dos rios e riachos (FIGURA 3D e E, PAVANELLI, 1999). O corpo apresenta a face ventral plana, conhecidos no Brasil como “canivetes”, “charutos”, “cigarrilhas”, os quais habitam ambientes lóticos de riachos de cabeceira (PAVANELLI, 1999; PAVANELLI; BRITSKI, 2003).

FIGURA 4. Representantes da família Parodontidae

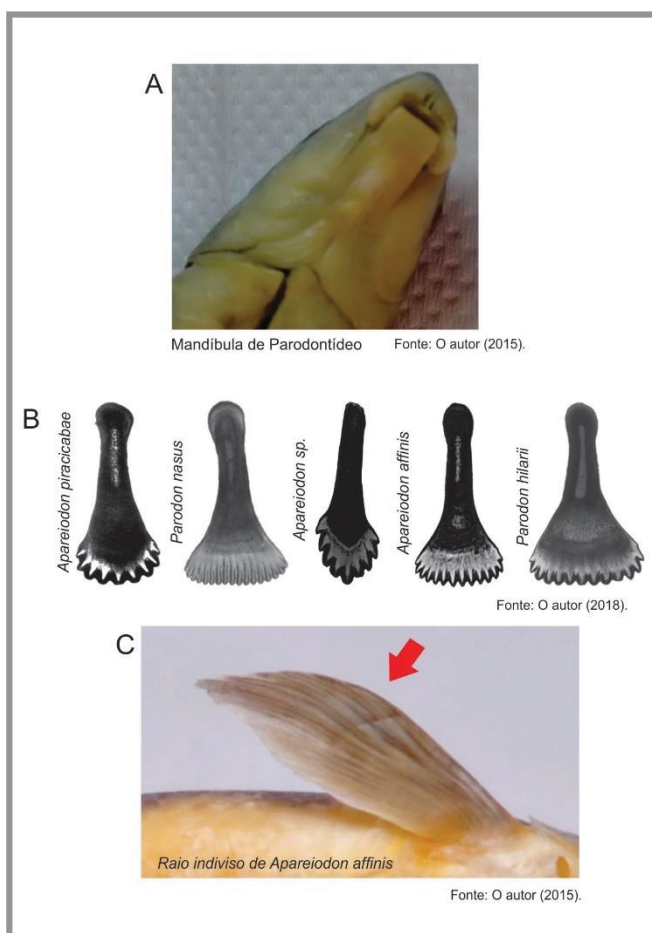


Legenda: Fotos dos representantes dos três gêneros da família Parodontidae mostrando o padrão de coloração (A-C). Foto de um representante do gênero *Apareiodon* evidenciando a face ventral plana (D) e as nadadeiras peitorais bem desenvolvidas (E). FONTE: O Autor (2015, 2018); PAVANELLI (2007).



Os representantes da família Parodontidae são diferenciados dos demais Characiformes pela ausência de lábio superior; dentes pré-maxilares frouxamente ligados ao osso pré-maxilar e dentes mandibulares, quando presentes, estão localizados na região lateral da boca e direcionados para dentro da cavidade bucal. (PAVANELLI, 1999). A mandíbula tem formato de uma pá de bordo reto, desprovidas de dentes na borda anterior (FIGURA 4A) e, a pré-maxila está rodeada por uma série de pequenos dentes com cúspides arredondados (FIGURA 4B). O número de cúspides de dentes pré-maxilares pode ser um caráter para o diagnóstico de espécies entre *Parodon* e *Apareiodon*, mas não para espécies *Saccodon*, que são caracterizados pelo grande polimorfismo dentário (GARAVELLO, 1977).

Os três gêneros de Parodontidae têm sido caracterizados pela variação de dois caracteres: número de raios indivisos nas nadadeiras peitorais - um raio em *Parodon* e *Apareiodon* e dois em *Saccodon* (FIGURA 4C) - e ausência de dentes na região lateral da mandíbula (*Apareiodon* e *Saccodon*) (PAVANELLI; BRITSKI, 2003; PAVANELLI, 2007). Não obstante, o padrão de coloração pode também ser utilizado para diferenciá-los (TRAVASSOS, 1955; ROBERTS, 1974; PAVANELLI; BRITSKI, 2003; PAVANELLI, 2007; INGENITO, 2008). No entanto, seus representantes possuem relações filogenéticas controversas por apresentarem poucos e sutis caracteres morfológicos confiáveis para sustentar as monofilias dos gêneros *Parodon* e *Apareiodon* (PAVANELLI; BRITSKI 2003; INGENITO 2008).

FIGURA 5. Caracteres morfológicos de exemplar do gênero *Apareiodon*

Legenda: Foto de um representante do gênero *Apareiodon* evidenciando a mandíbula (A), Dentes dos gêneros *Apareiodon* e *Parodon* (B) e Raio indiviso da nadadeira peitoral (C). FONTE: O Autor (2015, 2018).

## 2.5.2 Citogenética clássica e molecular

Na família Parodontidae os estudos citogenéticos têm mostrado um número diploide conservado de 54 cromossomos, sendo na maioria das espécies compostos basicamente por meta-submetacêntricos (m-sm), sendo os cromossomos subtelocêntricos (st) ausentes ou variando entre um a três pares entre as espécies (MOREIRA FILHO; BERTOLLO, GALETTI JUNIOR, 1993; BELLAFRONTE et al., 2011). No entanto, uma exceção ocorre em populações de *Apareiodon affinis* do baixo rio Paraná (Argentina), rio Paraguai, rio Uruguai e rio Cuiabá, onde um grande número de cromossomos acrocêntricos são observados (JORGE; MOREIRA FILHO, 2000; 2004; NASCIMENTO et al., 2018). Ainda, em *Apareiodon piracicabae* do rio Passa Cinco – SP foram diagnosticadas variações do número cromossômico 54 devido à

presença de cromossomos B (FALCÃO; MOREIRA FILHO; BERTOLLO, 1984). O gênero *Saccodon* até o momento não apresenta estudos citogenéticos.

Embora as espécies pertencentes à família apresentem um número diploide conservado, uma série de características cromossômicas e genômicas não são compartilhadas por todos seus representantes, tais como o sistema de cromossomos sexuais heteromórficos, polimorfismo estrutural dos cromossomos, cromossomos supranumerários, variação inter e intraespecífica no número e posição das regiões organizadoras de nucléolo (RONs) nos cromossomos e as diferenças interespecíficas na distribuição famílias de DNA satélite, número e localização de sítios de rDNA 18S e 5S, heterocromatina, localização e número de sítios do DNA satélite pPh2004, fração repetitiva WAp e acúmulo de TEs (MOREIRA FILHO; BERTOLLO, GALETTI JUNIOR, 1980, 1984, 1993; JESUS; MOREIRA FILHO, 2000a, 2000b; JORGE; MOREIRA FILHO, 2000, 2004; VICENTE; JESUS; MOREIRA FILHO, 2001; VICENTE et al., 2003; CENTOFANTE; BERTOLLO; MOREIRA FILHO, 2002; BELLAFRONTE; MARGARIDO; MOREIRA-FILHO, 2005; BELLAFRONTE et al., 2009, 2011, 2012; ROSA et al., 2006; VICARI et al., 2006; SCHEMBERGER et al., 2011, 2014, 2016; ZIEMNICZAK et al., 2014; NASCIMENTO et al., 2018).

Entre representantes de *Apareiodon*, os sítios de DNA ribossomais 45S, detectados pela técnica de impregnação de Nitrato de Prata (Ag-RONs) e sonda de DNA ribossomal (rDNA) 18S estão localizados na região terminal do braço longo de um par cromossômico subtelocêntrico grande (MOREIRA FILHO; BERTOLLO; GALETTI JUNIOR, 1984, 1985; JESUS; MOREIRA FILHO, 2000a; ROSA et al., 2006; VICARI et al., 2006; BELLAFRONTE et al., 2009, 2011), e presença de sítios adicionais em *Apareiodon ibitiensis*, *Apareiodon vittatus* e *Apareiodon* sp. – rio Aripuanã (BELLAFRONTE et al., 2009, 2011; SANTOS et al., 2019, em fase de pré-publicação). *Apareiodon hasemani* apresentou uma exceção onde o rDNA 45S foi localizado em pares meta/submetacêntricos (BELLAFRONTE et al., 2012).

Por sua vez, *A. piracicabae* mostra polimorfismo intra e interindividual relacionada com a ocorrência de RONs duplicadas em um par de cromossomos subtelocêntricos (MOREIRA FILHO; BERTOLLO; GALETTI JUNIOR, 1984). No gênero *Parodon*, as RONs têm se mostrado como caracteres cromossômicos espécie-específicos localizados em metacêntricos e/ou submetacêntricos (JESUS; MOREIRA FILHO, 2000b; VICENTE; JESUS; MOREIRA FILHO, 2001; CENTOFANTE; BERTOLLO; GALETTI JUNIOR, 2002; BELLAFRONTE et al., 2005, 2011), sugerindo

que em Parodontidae, a movimentação de sítios de rDNA ocorreram na diversificação cariotípica do grupo (BELLAFRONTTE et al., 2011).

No que se refere a localização do rDNA 5S, a região proximal de um par submetacêntrico foi verificada entre os parodontídeos estudados, característica que é homóloga a encontrada na família Anostomidae (CENTOFANTE; BERTOLLO; GALETTI JUNIOR, 2002; VICARI et al., 2006; BELLAFRONTTE et al., 2009, 2011; SCHEMBERGER et al., 2011), a exceção de *A. hasemani*, que apresenta rDNA 5S em posição subterminal de um par submetacêntrico. Vicente; Jesus; Moreira-Filho (2001) detectaram um maior e um menor conjunto de rDNA 5S em pares cromossômicos distintos de *Parodon hilarii*, *Parodon nasus* citado como *Parodon tortuosos* e *Parodon pongoensis*. A espécie *P. nasus* é a única a apresentar sintenia dos rDNAs 5S e 18S, característica apomórfica originada provavelmente por um evento de translocação cromossômica (BELLAFRONTTE et al., 2011).

## 2.5.2 Estudos citogenéticos e moleculares envolvendo os cromossomos sexuais

A família Parodontidae apresenta sistemas de cromossomos sexuais diversificados, com a ocorrência de espécies sem cromossomos sexuais heteromórficos, com proto-cromossomos sexuais, sistema de cromossomos sexuais simples (ZZ/ZW) ou múltiplo (ZZ/ZW<sub>1</sub>W<sub>2</sub>) (SCHEMBERGER et al., 2011; BELLAFRONTTE et al., 2012).

A origem do sistema de cromossomos sexuais tipo ZZ/ZW de Parodontidae ocorreu por uma inversão paracêntrica de um sítio terminal detentor da fração repetitiva denominada WAp para a região proximal do braço curto de um par cromossômico metacêntrico e, posterior amplificação desta sequência levando a diferenciação do cromossomo sexual W na maioria das espécies (SCHEMBERGER et al., 2011).

Para a origem do sistema múltiplo, MOREIRA-FILHO et al. (1993) propuseram a origem de sistemas múltiplos a partir de sistemas simples onde o W ancestral teria sofrido rearranjos do tipo fissão cêntrica, seguido de inversão pericêntrica originando os cromossomos W<sub>1</sub> e W<sub>2</sub>. Alguns estudos posteriores elucidaram mecanismos envolvendo a diferenciação dos cromossomos sexuais neste grupo, a começar pela localização da heterocromatina na identificação do sistema de cromossomos sexuais do sistema ZZ/ZW. Os blocos de heterocromatina em Parodontidae foram localizados

nas regiões centroméricas e pericentroméricas, com evidentes marcações terminais em todas as espécies de Parodontidae analisadas (MOREIRA FILHO; BERTOLLO; GALETTI JUNIOR, 1984,1993; JESUS; BERTOLLO; MOREIRA-FILHO, 1999; JESUS; MOREIRA-FILHO, 2000a; JORGE; MOREIRA FILHO, 2004; ROSA et al., 2006; VICARI et al., 2006; BELLAFRONTE et al., 2009, 2012).

Nas espécies que possuem cromossomos sexuais do tipo ZZ/ZW, a heterocromatina constitutiva no cromossomo Z apresenta pequenos blocos heterocromáticos, sendo um proximal e outro terminal no braço curto. Entretanto, o cromossomo W sofreu um processo de heterocromatização e acúmulo de sequências repetitivas na diferenciação sexual. A heterocromatização e o acúmulo de sequências repetitivas do W ocorreram a partir de uma pequena região heterocromática proximal do braço curto presente no cromossomo Z, a qual, no cromossomo W diferenciado é maior e corresponde ao braço longo. Esta hipótese de diferenciação é utilizada para explicar a origem evolutiva do sistema ZZ/ZW presente em *P. hilarii*, *P. moreirai* (citado como *Parodon* sp.), *A. vladii* (citado como *Apareiodon* sp.), *Apareiodon* sp. (localizado no rio Verde) e *A. ibitiensis*. (CENTOFANTE; BERTOLLO; MOREIRA FILHO, 2002; VICENTE et al., 2003; ROSA et al., 2006; VICARI et al., 2006; BELLAFRONTE et al., 2009; SCHEMBERGER et al., 2011). A exceção ocorre em *A. hasemani*, onde o acúmulo de heterocromatina no cromossomo W ainda é pequeno, o que deixa a região de diferenciação do cromossomo W localizada no braço curto (BELLAFRONTE et al., 2012).

Estudos de localização *in situ* a partir do DNA satélite isolado de *Parodon hilarii* pelo método de restrição genômica do DNA, denominado DNA satélite pPh2004 evidenciaram 16 sítios em autossomos, além de um sítio na região terminal do braço longo do cromossomo Z e um sítio na região terminal do braço curto do cromossomo W de *P. hilarii* (VICENTE et al., 2003). O satélite pPh2004 também foi localizado em cromossomos de *Parodon moreirai* (CENTOFANTE; BERTOLLO, MOREIRA FILHO, 2002) onde foram encontrados os mesmos sítios encontrados nos cromossomos Z e W de *P. hilarii*, além de dois pequenos sítios adicionais nos autossomos.

O mapeamento *in situ* do pPh2004 em alguns representantes de Parodontidae evidenciou diferenças na localização e número de sítios para *P. nasus*, *P. pongoensis*, *P. moreirai*, *P. hilarii* e *A. affinis* (BELLAFRONTE et al., 2011; SCHEMBERGER et al., 2011; NASCIMENTO et al., 2018). As espécies *A. ibitiensis*, *A. affinis* (do Rio Uruguai), *A. vladii*, *A. piracicabae* não apresentaram marcações deste DNA satélite

(BELLAFRONTTE et al., 2011; SCHEMBERGER et al., 2011; NASCIMENTO et al., 2018). Assim, Bellafronte et al. (2011) propuseram que as espécies que possuem o DNA satélite pPh2004 divergiram mais recentemente e são mais relacionadas. Ainda para populações de *A. affinis* dos rios Paraguai e Cuiabá inúmeros sítios pPh2004 foram detectados (NASCIMENTO et al., 2018).

Em outro estudo, uma fração repetitiva de DNA foi obtida dos cromossomos sexuais de Parodontidae pela microdissecção do cromossomo W de *Apareiodon* sp. (VICARI et al., 2010). A sonda foi denominada WAp e utilizada na hibridização *in situ* fluorescente (FISH) onde detectou sítios em regiões terminais de alguns cromossomos do cariótipo de todas as espécies de Parodontidae estudadas (*Apareiodon* sp., *A. ibitiensis*, *A. vladii*, *A. piracicabae*, *A. vittatus*, *A. affinis*, *P. hilarii*, *P. moreirai*, *P. nasus* e *P. pongoensis*) além de marcações nos cromossomos sexuais em *A. affinis* (ZZ/ZW<sub>1</sub>W<sub>2</sub>) e; em todas as espécies com sistema de cromossomos sexuais do tipo ZZ/ZW (SCHEMBERGER et al., 2011).

As marcações da sonda WAp nos cromossomos Z e W das espécies com sistema ZZ/ZW corroboram com a hipótese de que houve um acúmulo de sequências repetitivas compartilhadas entre as espécies para a derivação do cromossomo W (CENTOFANTE; BERTOLLO; MOREIRA FILHO, 2002; VICENTE et al., 2003; ROSA et al., 2006; VICARI et al., 2006; BELLAFRONTTE et al., 2009, 2011, 2012; SCHEMBERGER et al., 2011, 2014). A sonda WAp também detectou marcações em regiões terminais de vários autossomos em todas as espécies analisadas, ratificando que esta sequência não é exclusiva dos cromossomos sexuais (SCHEMBERGER et al., 2011).

Assim, pode-se inferir uma possível origem dos cromossomos sexuais por um evento de inversão de sequências WAp de regiões terminais para a região proximal do braço curto de um par metacêntrico, originando um par de cromossomos proto-sexuais (presente nas espécies *P. nasus* e *P. pongoensis*) seguido de acúmulo destas sequências na origem do cromossomo W heteromórfico (SCHEMBERGER et al., 2011).

A localização *in situ* da sonda WAp em conjunto com o DNA satélite pPh2004 em Parodontidae possibilitou verificar a ocorrência de espécies sem cromossomos sexuais heteromórficos (*Apareiodon piracicabae* e *Apareiodon vittatus*); com proto cromossomos sexuais e presença de DNA satélite pPh2004 (*Parodon nasus*, *Parodon pongoensis*, *Apareiodon argenteus* e *Apareiodon davisii*); com sistema de



cromossomos sexuais heteromórficos (*Apareiodon* sp. – rio Verde, *Apareiodon ibitiensis*, *Apareiodon vladii* e *Apareiodon hasemani*); com sistema de cromossomos sexuais heteromórficos e presença de DNA satélite pPh2004 (*Apareiodon* sp. – rio Aripuanã, *Parodon moreirai* e *Parodon hilarii*); além da espécie *Apareiodon affinis* (sistema do Alto Rio Paraná), com sistema de cromossomos sexuais múltiplos e DNA satélite pPh2004 (SCHEMBERGER et al., 2011; BELLAFRONTE et al., 2012; TRALDI et al., 2016; NASCIMENTO et al., 2018; SANTOS et al., 2019, em fase de pré-publicação).

De modo semelhante, a utilização das sondas WAp e pPh2004 em *A. affinis* foi utilizada para explicar a origem do sistema sexual ZZ/ZW<sub>1</sub>W<sub>2</sub>, a qual considera a ocorrência de rearranjos cromossômicos do tipo translocação autossomo proto cromossomos sexuais portadores de DNA satélite pPh2004 (SCHEMBERGER et al., 2011). Posteriormente, rearranjos do tipo fissão cêntrica e inversão pericêntrica originaram os cromossomos W<sub>1</sub> e W<sub>2</sub> conforme proposto por MOREIRA-FILHO; BERTOLLO; GALETTI JUNIOR (1980).

Estudos de composição de sequências repetitivas presentes no cromossomo W de Parodontidae descreveram que a região pseudoautossômica (RPA) dos cromossomos sexuais é composta pelos braços longo do cromossomo Z e, curto do cromossomo W, as quais possuem o DNA satélite pPh2004 no gênero *Parodon* (SCHEMBERGER et al., 2014). Na região específica do cromossomo W (REW) de Parodontidae foi possível verificar o acúmulo de sequências satélites e de elementos transponíveis (TEs), as quais não são exclusivas dos sexuais (SCHEMBERGER et al., 2014). A localização proximal da REW é composta pelo DNA satélite (sat1WP). No terminal da REW foi visualizado um extensivo acúmulo do TE tipo Helitron e; em toda sua extensão, a REW é invadida por TEs degenerados de diferentes superfamílias (SCHEMBERGER et al., 2014). Ainda, uma expansão do microsatélite (GATA)<sub>n</sub> (ZIEMNICZAK et al., 2014), elemento transponível Tc1-Mariner e o DNA satélite (sat1WAP) (SCHEMBERGER et al., 2016) foram confirmados nos dados genômicos e localizados na REW de *Apareiodon* sp. Assim, somados a inúmeras expansões microsatélites (CA)<sub>n</sub>, (CAA)<sub>n</sub>, (CAT)<sub>n</sub> e (CG)<sub>n</sub> e aos TEs hAT, EnSpm e Helitron é sugerido que estes elementos repetitivos atuaram na diferenciação do cromossomo W (SCHEMBERGER et al., submetido à publicação).

Portanto, acúmulos de sequências repetitivas *in tandem* e dispersas no cromossomo W podem ter contribuído para a diferenciação molecular e a erosão da

atividade gênica do cromossomo W de *Apareiodon* sp., com possível papel na formação domínios na cromatina e no controle da expressão dos prováveis genes de determinação sexual presente nos cromossomos sexuais.



### 3 OBJETIVOS

#### 3.1 GERAL

Este estudo teve como objetivo avaliar o papel de DNAs repetitivos com foco na diversidade cariotípica, evolução dos cromossomos sexuais e origem de DNAs satélites na família Parodontidae.

#### 3.2 ESPECÍFICOS

- Identificar os elementos repetitivos a partir do sequenciamento em larga escala de *Apareiodon* sp.;
- Recuperar as sequências de diferentes elementos de transposição;
- Caracterizar a estrutura molecular de elementos transponíveis Gypsy e En/Spm e verificar sua atividade no genoma;
- Localizar por hibridização *in situ* fluorescente os elementos repetitivos (Transposon de DNA, Retrotransposon e DNA microssatélites) em representantes da família Parodontidae;
- Investigar a atuação de elementos repetitivos na origem e evolução dos cromossomos sexuais;
- Avaliar a possível participação dos elementos transponíveis na origem de DNA satélite em *Apareiodon* sp.
- Estudar a dinâmica evolutiva dos elementos repetitivos na família Parodontidae.

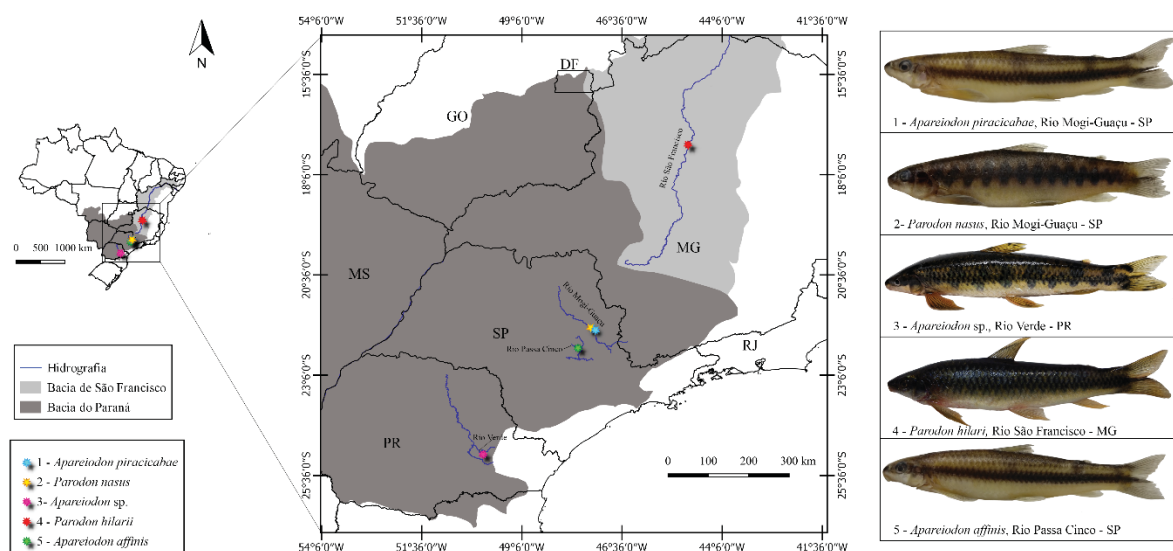
## 4 MATERIAL E MÉTODOS

### 4.1 MATERIAL

O material biológico utilizado neste estudo foi obtido de espécimes de peixes da família Parodontidae coletados nos rios Mogi-Guaçu – SP, Verde – PR e Passa Cinco – SP pertencentes a bacias hidrográficas do Alto Paraná e do rio São Francisco – MG pertencente a bacia do São Francisco (FIGURA 5, TABELA 1).

As espécies utilizadas foram *Apareiodon piracicabae* (16 machos e 12 fêmeas), *Parodon nasus* (11 machos e 9 fêmeas), *Apareiodon* sp. (28 machos e 24 fêmeas), *Parodon hilarii* (24 machos e 16 fêmeas) e *Apareiodon affinis* (20 machos e 14 fêmeas) (FIGURA 5, TABELA 1). Os exemplares utilizados neste estudo foram obtidos via coleta de peixes na natureza com licença do Ministério do Meio Ambiente e ICMBIO (Licença permanente para coleta de material zoológico MMA/ICMBIO/SISBIO: 15117-1) (ANEXO 1). Os procedimentos estão de acordo com o Comitê de Ética do Uso de Animais da Universidade Estadual de Ponta Grossa (Protocolo: 11/2017) (ANEXO 2).

FIGURA 6. Pontos de coleta dos exemplares da família Parodontidae



Legenda: Mapa do Brasil com destaque para as bacias hidrográficas do Paraná e São Francisco mostrando os rios e os pontos de coletas. À direita as espécies estudadas numeradas de acordo com a sua localização nas bacias. FONTE: O Autor (2019)

**TABELA 1.** Espécies da família Parodontidae estudadas, localidade, distribuição em bacias hidrográficas e número de espécies.

Espécies	Rio, Estado	Bacia Hidrográfica	Número de ♂	Número de ♀	GPS
<i>A. piracicabae</i>	Mogi-Guaçu - SP	Alto Paraná	16	12	-21°55'38.6" e -47°21'58.4"
<i>P. nasus</i>	Mogi Guaçu - SP	Alto Paraná	11	9	-21°54'58" e -47°22'58"
<i>Apareiodon</i> sp.	Verde - PR	Alto Paraná	28	24	-25°04'35" e -50°04'03"
<i>P. hilarii</i>	São Francisco - MG	São Francisco	24	16	-17°21'17" e -44°57'15"
<i>A. affinis</i>	Passa Cinco - SP	Alto Paraná	20	14	-22°25'26" e -47°41'56"

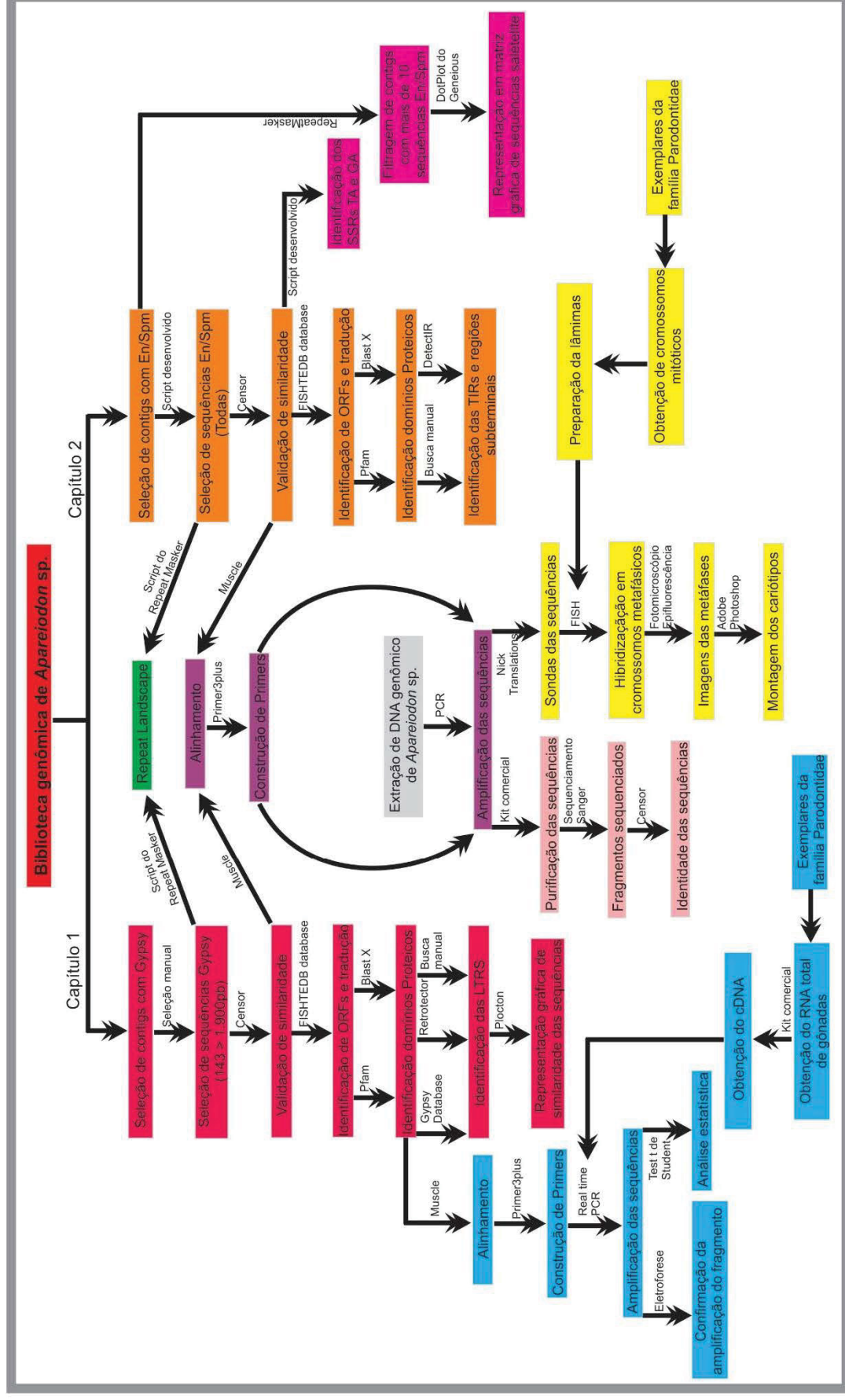
FONTE: O Autor (2019)

## 4.2 METODOLOGIA

Este estudo foi baseado nos dados do sequenciamento do genoma completo de *Apareiodon* sp. A fração repetitiva do genoma foi anotada utilizando o software RepeatMasker (SMIT et al., 2013-2015) e genoma de referência *Danio rerio* por Schemberger et al. (submetido à publicação).

As etapas desenvolvidas durante este estudo estão representadas no fluxograma na FIGURA 6.

FIGURA 7. Fluxograma representativo das etapas desenvolvidas durante este estudo. Cada cor representa um tipo de análise.



FONTE: O Autor (2019).

#### 4.2.1 Caracterização molecular retrotransposon Gypsy e Transposon de DNA En/Spm

Para a caracterização molecular do Gypsy e En/Spm foram adotadas diferentes abordagens. Os softwares, script e estratégias utilizadas nesta etapa estão descritos em material de métodos de cada capítulo correspondente.

#### 4.2.2 Identificação e análise de DNA satélites

A busca e identificação de DNA microssatélites foi realizada com os dados do RepeatMasker (SMIT et al., 2013-2015) e um script de linguagem em php foi desenvolvido para auxiliar nesta pesquisa. Posteriormente foi realizada uma análise em matriz gráfica Dotplot do software Geneious 4.8.5 (DRUMMOND et al., 2009) com mecanismo de comparação baseado na ferramenta Dottup.

#### 4.2.3 Construção de primers

As sequências de *Apareiodon* sp. com similaridade aos elementos Gypsy e En/Spm foram recuperadas. As sequências com menor degeneração foram submetidas ao alinhamento MUSCLE (EDGAR, 2004) no software Geneious 4.8.5 (DRUMMOND et al., 2009) e as regiões com menor degeneração foram utilizadas para a construção de primers no software Primer3plus (UNTERGASSER; NIJVEEN, 2007; ROZEN; SKALETISKY, 2000). Foram sintetizados seis conjuntos de primers para validação RTE Gypsy e um conjunto para o transposon En/Spm (ANEXO 3).

#### 4.2.4 Extração de DNA genômico

O DNA genômico foi obtido a partir do fígado dos exemplares de *Apareiodon* sp. utilizando brometo de cetil trimetilamônio (CTAB) de acordo com o protocolo adaptado de Doyle; Doyle (1990) (ANEXO 4).

#### 4.2.5 Amplificação das sequências Gypsy e En/Spm para sequenciamento e localização *in situ*

Para a amplificação das sequências Gypsy realizamos uma reação em cadeia da polimerase (PCR) de longa extensão utilizando o primer Gypsy P1-Fw e Gypsy P6-Rv com o Kit Long Amplifications Taq DNA polymerase da Ludwig Biotec.

Posteriormente realizamos PCRs para cada conjunto de primer Gypsy, em duplicata, sendo uma utilizando o produto da long PCR como molde e outra com DNA genômico.

Os fragmentos do retrotransposon Gypsy e transposon En/Spm foram amplificados por PCR. Os parâmetros de amplificação e tamanho do fragmento esperado foram de acordo com a tabela disponível no ANEXO 5.

Os produtos da PCR dos conjuntos de Primer Gypsy\_P3 e En/Spm foram utilizados para marcação de sonda e posterior uso na FISH.

#### 4.2.6 Purificação e sequenciamento dos fragmentos Gypsy e En/Spm

Os produtos de PCR (duplicatas) do elemento Gypsy e En/Spm foram purificados utilizando o kit PCR DNA and Gel Band Purification (GE Healthcare), conforme protocolo do fabricante (ANEXO 6). Posteriormente foram submetidos ao sequenciamento automático de DNA utilizando sequenciador de DNA ABI-PRISM Genetic Analyzer (Applied Biosystems) e submetido ao software online CENSOR para a confirmação de identidade do segmento (KOHANY et al., 2006).

#### 4.2.7 Marcação de sondas

As sequências do transposon de DNA En/Spm, retrotransposon Gypsy resultante da PCR foram marcadas com digoxigenina-11-dUTP utilizando o Kit Dig-Nick Translation Mix (Roche Applied Science) para obtenção da sonda (ANEXO 7).

As sequências microssatélites TA<sub>15</sub> e GA<sub>15</sub> foram marcadas com biotina na extremidade 5' durante sua síntese (Sigma-Aldrich, St. Louis, Missouri, EUA).

#### 4.2.8 Preparações mitóticas

Os espécimes coletados foram mantidos em aquários aerados e submetidos a um processo inflamatório para induzir o número de células em divisão conforme descrito por Cavallini; Bertollo (1988) (ANEXO 8). A obtenção de cromossomos

mitóticos foi realizada seguindo a técnica de preparações diretas (ANEXO 9) (BERTOLLO; CIOFFI; MOREIRA FILHO, 2015).

#### 4.2.9 Preparação das lâminas

Para confecção de lâminas cerca de 100 microlitros da suspensão celular com material cromossômico foram gotejados sobre as lâminas com auxílio de uma pipeta Pasteur. Foram utilizadas lâminas limpas, levemente inclinada, com uma película d'água a 55°C para permitir o espalhamento e a fixação do material sobre a lâmina. O material foi então corado com solução de Giemsa a 5% em tampão fosfato pH=6,8 durante 6-8 minutos (ANEXO 9) (BERTOLLO; CIOFFI; MOREIRA FILHO, 2015).

#### 4.2.10 Hibridação in situ fluorescente (FISH)

O procedimento geral da FISH foi conforme descrito por Pinkel; Straume; Gray (1986), (ANEXO 10). Foram utilizadas sondas de En/Spm, Gypsy e DNAs microssatélites GA<sub>15</sub> e TA<sub>15</sub> para a localização dessas sequências nos cromossomos metafásicos das espécies *Apareiodon piracicabae* (sem cromossomos sexuais heteromórficos), *Parodon nasus* (proto-cromossomos sexuais) *Apareiodon* sp. (ZZ/ZW), *Parodon hilarii* (ZZ/ZW) e *Apareiodon affinis* (ZZ/ZW<sub>1</sub>W<sub>2</sub>).

#### 4.2.11 Análises dos resultados e montagem dos cariótipos

As preparações cromossômicas com hibridização in situ fluorescente foram analisadas em microscópio de campo claro e epifluorescência Zeiss Axio Lab 1 acoplado ao software Zen Lite em Câmera Zeiss AxioCam ICM1, resolução de 1.4 megapixels (Carl Zeiss®).

As melhores metáfases foram selecionadas para organizar os cariótipos, onde os cromossomos foram recortados usando o software Adobe Photoshop, versão CS6, e organizados em pares de homólogos em classes e ordem decrescente de tamanho. A classificação dos cromossomos em metacêntricos (m), submetacêntricos (sm), subtelocêntricos (st) e acrocêntricos (a) foi realizada conforme proposto por Levan; Fredga; Sandberg (1964).

#### 4.2.12 Extração de RNA e análise de expressão

Para a análise de expressão foram construídos dois conjuntos de primers para os domínios proteicos mais conservados do RTE-Gypsy: Transcriptase reversa e Integrase e um terceiro conjunto para o domínio proteico PNMA (ANEXO 11).

O tecido gonadal de *Apareiodon* sp. foi extraído e armazenado em RNAlater® em temperatura de - 80°. Posteriormente foi realizada a extração do RNA total de usando TRIZOL (Invitrogen), de acordo com as instruções do fabricante (ANEXO 12).

O RNA total foi submetido a síntese de DNA complementar utilizando o Kit da GE Healthcare First Strand cDNA Synthesis, seguindo as instruções do fabricante (ANEXO 13). Os cDNAs foram amplificados em duplicatas por PCRs utilizando o SYBR Green master mix (Roche). A  $\beta$ -actina foi utilizada como controle interno.

Os reagentes utilizados, parâmetros de amplificação, análise dos fragmentos e análise da gráfica estão descritos no capítulo 1.



## **5 RESULTADOS**

Os resultados obtidos neste estudo estão apresentados em forma de artigos científicos divididos em 2 capítulos:

### **5.1 CAPÍTULO 1**

**ANCIENT INVASION PHASE, MOLECULAR DEGENERATION AND PROBABLE CO-OPTATED PNMA DOMAIN OF RETROTRANSPOSON GYPSY IN THE GENOME OF A NEOTROPICAL PARODONTIDAE FISH**

### **5.2 CAPÍTULO 2**

**ELEMENTO MÓVEL EN/SPM SENESCENTE EM UM GENOMA DE PEIXE NEOTROPICAL: GÊNESE DE REPETIÇÕES IN TANDEM E MECANISMOS MOLECULARES NA DIFERENCIAÇÃO DOS CROMOSSOMOS SEXUAIS ZW**

## 5.1 CAPÍTULO 1

### ANCIENT INVASION PHASE, MOLECULAR DEGENERATION AND PROBABLE CO-OPTATED PNMA OF RETROTRANSPOSON GYPSY IN THE GENOME OF A NEOTROPICAL PARODONTIDAE FISH

#### ABSTRACT

Retrotransposable elements (RTEs) are considered abundant in fish genomes. Parodontidae species show relatively small genomes (~ 1.2 Gb). Our data showed that RTEs correspond to 27% of the repetitive fraction and L1, L2, Rex-Babar and Gypsy are the most abundant ones in the genome of *Apareiodon* sp. We have assembled the genome of *Apareiodon* sp. based in Illumina sequencing and recovered contigs that showed similarity to the RTE Gypsy. Additionally, Gypsy sequences were amplified by PCR and resequenced. Subsequently, an analysis of the molecular structure of the LTR-5', LTR-3' and open reading frames (ORFs) for *Gag* and *Pol* polyproteins was performed. Gypsy sequences in the genome of *Apareiodon* sp. were degenerated, with most copies showing 20 – 42% nucleotide substitutions in a landscape analysis. However, some sequences with the protein domains present in *Gag* and *Pol* were detected and analyzed. Also, upstream to the *Gag* sequence, a region of similarity to the PNMA (Paraneoplastic Ma antigens, Gag-derived protein) present in mammals was detected. The presence of intact Gypsy protein domains has been discussed, which may provide material for origin of new molecular structures, like PNMA proteic domain or may be kept neutral in this genome. The *in situ* chromosomal location of the Gypsy sequences in Parodontidae species showed scattered signals in the chromosomes and did not demonstrate accumulations in the W heteromorphic chromosome. Our data evidences molecular degeneration of the RTE-LTR Gypsy in Parodontidae and its apparent lack of mobilization since the diversification of the group.

**Keywords:** genomic studies, *in situ* localization, molecular deterioration, neutral and co-opted retrotransposons, repetitive elements

## 1 INTRODUCTION

The investigation of the repetitive fraction of eukaryotic genomes has addressed important contributions in understanding genomic structure and function (Arkhipova 2017). The recent advances of new genomic technologies including next generation sequencing (NGS) and development of bioinformatics tools increases the power to analyse large-scale genomic and transcriptional data (Macas et al. 2011; Palazzo and Gregory 2014; Valente et al. 2014; Amselem et al. 2019). Such new technologies have increased our capability to investigate structural and functional issues of unknown fractions of genomes, thus enabling studies at a level that was previously never considered possible (Valente et al. 2017).

The transposable elements (TEs) correspond to an important fraction of the repetitive DNAs and are classified into two classes, each one organized into orders, superfamilies, families and subfamilies (Wicker et al. 2007). Class I is represented by retrotransposons (RTEs), which move by reverse transcription through an intermediate RNA, and class II, represented by DNA transposons, transpose into the genomes via their DNA molecules (Finnegan 1989; Wicker et al. 2007). TEs can also be classified as autonomous and non-autonomous elements, regardless of their class (Wicker et al. 2007). Autonomous TEs encode all functional and necessary enzymes for transposition, while non-autonomous ones have truncated/degenerate copies, although they can be transposed using the enzymes produced by autonomous elements (Wicker et al. 2007; Hua-Van et al. 2011).

The RTEs can be divided into five major orders according to the molecular structure and phylogeny of reverse transcriptase: (i) long terminal repeat (LTR) retrotransposons, (ii) *Dictyostelium* intermediate repeating sequence (DIRS), (iii) Penelope (PLE), (iv) long interspersed nuclear elements (LINEs), and (v) short interspersed nuclear elements (SINEs) (Malik et al. 1999; Wicker et al. 2007; Eickbush and Jamburuthugoda 2008). RTE-LTRs are widely distributed in eukaryotic genomes (Eickbush and Malik 2002) and are subdivided into five superfamilies: Copy, Gypsy, Bel-Pao, Retrovirus and ERV (Wicker et al. 2007). Each RTE-LTR is a functional unity composed by two modules, with distinct structure and function: (i) two LTRs in their flanking regions which have 5'-TG-3' start and 5'-CA-3' and flanks the coding region (the LTRs range from a few hundred base pairs to more than 5 kb and have transcriptional regulatory sequences); and (ii) an inner region coding for *Gag* (group

specific antigen) and *Pol* (Poliprotein) genes required for retrotransposition, ranging from a few hundred base pairs to 25 kb (Peterson-Burch and Voytas 2002; Benachenhou et al. 2009).

The superfamily Gypsy, as well as the other RTE-LTRs, present the poliproteins *Gag* and *Pol*. *Gag* contains three proteic domains: matrix (MA), capsid (CA) and nucleocapsid (NC). In addition, *Pol* has four protein domains: aspartic proteinase (AP), reverse transcriptase (RT), RNase H (RH), integrase (INT) and in some groups of plants, fungi and vertebrates *Pol* can also present the Chromo (CHR) domain. *Gag* and *Pol* genes generally overlap in different open reading frames (ORFs) however, there are numerous examples where *Gag* and *Pol* are fused to a single ORF (Wicker et al. 2007; Eickbush and Jamburuthugoda 2008; Rawlings et al. 2008).

Genomic studies have shown that TEs, after the invasion phase, tend to be silenced and suffer molecular deterioration (Fernández-Medina et al. 2012; Lowe and Haussler, 2012; Schemberger et al. 2016). The movement of TEs in the genomes generate mutations and, as consequences, they can: (i) inactivate genes, (ii) become involved in chromosomal rearrangements, (iii) remain inactive as neutral DNAs in the genomes or (iv) provide events of molecular evolution (Kidwell and Lisch 2001; Feschotte 2008; Herpin et al. 2010; Kapusta et al. 2013; Glugoski et al. 2018). The most important feature of TEs is related to their ability to generate evolutionary novelties. TE sequences undergo a so-called molecular co-option process (Miller et al. 1997) that may cause the origin of new regulatory sequences, new splice and polyadenylation sites, micro RNAs, new genes, among others.

Although TEs have been extensively investigated among vertebrates (Warren et al. 2015), we have a gap of information for most fish groups. The molecular/functional characterization of TEs is scarce in Neotropical fishes (Schemberger et al. 2016) and the high diversity of fishes in the Neotropics calls the attention to the emergence of new models to advance in understanding TEs. Parodontidae is a small group of Neotropical fish with striking characteristics related to occurrence of species without heteromorphic sex chromosomes, with sexual proto-chromosomes, and others with heteromorphic sex chromosome systems ZZ/ZW or ZZ/ZW<sub>1</sub>W<sub>2</sub> (Schemberger et al. 2011). Several studies have shown that the W sex chromosome of Parodontidae has a heterogeneous nature related to the composition of repetitive DNAs (Vicente et al. 2003; Vicari et al. 2010; Schemberger et al. 2011,

2014, 2016; Ziemniczak et al. 2014). These studies also showed that much of the karyotype diversification observed in the family is due to the accumulation and movement of repetitive DNAs (Bellafronte et al. 2011; Traldi et al. 2016; do Nascimento et al. 2018). In this way we have advanced over the characterization of the RTE-LTR Gypsy superfamily in the genome of *Apareiodon* sp., including additional analysis of molecular degeneration and chromosomal mapping in several species of the Parodontidae group. Our data, associated to previous TE studies, highlights paraodontids as an emerging model to investigate genome diversity and evolution.

## 2 MATERIALS AND METHODS

### 2.1 NEXT GENERATION SEQUENCING (NGS) AND GENOME ASSEMBLY

Liver tissue samples of *Apareiodon* sp., an undescribed species related to *A. ibitiensis* ( $2n = 54$  chromosomes and ZZ/ZW) were subject to genomic DNA extraction using the Kit Illustra tissue and cells genomic Prep Mini Spin (GE healthcare), following the manufacturer's protocol. Male and female genomic sequencing was performed by the Illumina HiSeq platform (paired-end 100-base reads). Reads were filtered using FASTX-Toolkit software to select reads showing  $\geq 90\%$  of sequenced bases with Phred score over 30. After filtering reads allowed  $\sim 40\times$  ( $\sim 50$  Gb) genome coverage. The Velvet software (Zerbino and Birney, 2008) was used to assemble the genome.

### 2.2 IDENTIFICATION, LANDSCAPE ANALYSIS AND AMPLIFICATION OF GYPSY SEQUENCES

The annotation of the repetitive fraction of the genome of *Apareiodon* sp. was performed with the software RepeatMasker (Smit et al. 2013-2015) using *Danio rerio* genome as reference and described in Schemberger et al. (submitted for publication). For Gypsy annotation it was used a custom repeat library obtained by RepeatModeler (Smit and Hubley 2008-2015) using default parameters. The custom library was merged with Repbase Update 20170127. The RepeatMasker was run with “slow” (-s), “library” (-lib) and “align” (-a) parameters. To summarize RepeatMasker results we used buildSummary.pl. Output files from RepeatMasker were also used as input for

create `calcdivergencefromalign.pl` and `repeatlandscape.pl`, in order to calculate Kimura divergence values and plot the repeat landscape.

The sequences identified as Gypsy were selected. The fullest Gypsy sequences obtained in the genomic annotation of *Aporeiodon* sp. were then submitted to the MUSCLE alignment (Edgar 2004) and regions of lower degeneration were used to construct primers using Primer3plus software (Untergasser and Nijveen 2007). Six sets of primers for polymerase chain reaction (PCR) validation of the longest Gypsy copy obtained in the genome of *Aporeiodon* sp. were synthesized (Supplementary Table 1).

Genomic DNA was obtained from the liver of the 14 specimens of *Aporeiodon* sp. using cetyl trimethyl ammonium bromide (CTAB) according to the protocol adapted from Doyle and Doyle (1990) to PCRs amplifications. In a first reaction a long-length PCR was performed with the Gypsy P1-Fw and Gypsy P6-Rv primers using 200 ng of genomic DNA, 0.2  $\mu$ M of each primer, 0.4 mM of deoxynucleotide triphosphates (dNTPs), 1  $\times$  Taq reaction buffer (200 mM Tris, pH 8.4, 500 mM KCl), 2 mM  $MgCl_2$  and 5 U of Long Amplification Taq DNA Polymerase (Ludwig Biotec). Amplification parameters were: initial denaturation 2 min at 95  $^{\circ}C$ , 35 three-step cycles (45 s at 94  $^{\circ}C$ , 45 s at 55  $^{\circ}C$  and 7 min at 68  $^{\circ}C$ ) and final extension of 5 min at 68  $^{\circ}C$ . Later, individual PCRs with the six primer sets were performed using the previous PCR product and also with genomic DNA as templates for the amplification and validation of the sequences. The mixing PCR consisted of: 40 ng of genomic DNA, 0.2  $\mu$ M of each primer, 0.4 mM of deoxynucleotide triphosphates (dNTPs), 1 $\times$  Taq reaction buffer (200 mM Tris, pH 8.4, 500 mM KCl), 2 mM  $MgCl_2$  and 1 U of Taq DNA polymerase (Invitrogen). Amplification parameters were: initial denaturation 5 min at 94  $^{\circ}C$ , 35 three step cycles (1 min at 94  $^{\circ}C$ , 45 s at 55  $^{\circ}C$  and 2 min at 72  $^{\circ}C$ ) and final extension of 7 min at 72  $^{\circ}C$ . PCR products were purified by the PCR DNA and Gel Band Purification Kit (GE Healthcare) and subjected to Sanger automatic DNA sequencing using the ABI-PRISM Genetic Analyzer (Applied Biosystems).

## 2.3 MOLECULAR CHARACTERIZATION OF GYPSY

The Gypsy sequences obtained from genomic annotation were submitted to the online CENSOR software for similarity validation (Kohany et al. 2006). We

identified 145 Gypsy sequences  $\geq 1,900$  bp in size that were selected and subjected to identification of open reading matrices (ORFs) and translation into amino acid sequences using the FishTEDB online database (Shao et al. 2018). The set of Gypsy sequences were then submitted to NCBI Blastx (Altschul et al. 1990) and the matching Pfam families (El-Gebali et al. 2019) for the identification of protein domains. Sequences lacking protein domains characteristic of Gypsy were excluded from the analysis. Subsequently, the intact sequences of each protein domain were aligned using MUSCLE (Edgar 2004) for the analysis of conserved regions.

For the recognition of the LTRs, contigs containing Gypsy sequences were submitted to analysis in the software Gypsy Database 2.0 (Llorens et al. 2011) and Retrorector online (Sperber et al. 2007) as well as manual search for similarity of consensus sequences U3, R and U5 present in LTRs, as described in Benachenhou et al. (2013). Gypsy DNA sequence alignments divided into the LTR-5', LTR-3', *Gag*, *Pol* and PNMA regions (Paraneoplastic Ma antigens, Gag-derived protein) were submitted to the EMBOSS Plotcon (<http://emboss.bioinformatics.nl/cgi-bin/emboss/plotcon>) to obtain a graphical representation of the similarities between the sequences and characterize their molecular deterioration level. The Wsize (Window Size) value for this study was 5.

## 2.4 RNA EXTRACTION AND TRANSCRIPTION ANALYSIS

Primers were synthesized to gene expression analysis in internal DNA sequences regions to Gypsy RT, INT and PNMA domains (Supplementary Table 2). Total RNA of *Apareiodon* sp. males and females gonad tissue was isolated using TRIZOL (Invitrogen), according to the manufacturer's instructions. The total RNA was submitted to First Strand cDNA Synthesis Kit (GE Healthcare), as recommended by the manufacturer's protocol. The cDNAs were amplified in conventional PCR using default parameters for each primer set (RT, INT, PNMA - Supplementary Table 2). The PCRs were submitted to electrophoresis in agarose gel 3 %.

## 2.5 IN SITU LOCATION

Chromosomal preparations of four species (Supplementary Table 3) belonging to Parodontidae: *Apareiodon piracicabae* (without sex heteromorphic chromosome),



*Apareiodon* sp. (ZZ/ZW), *Parodon hilarii* (ZZ/ZW) and *Apareiodon affinis* (ZZ/ZW<sub>1</sub>W<sub>2</sub>) were obtained according to the method described by Bertollo et al. (2015) and then subjected to fluorescence in situ hybridization (FISH), according to Pinkel et al. (1986). The procedures of this study are in agreement with the Committee of Ethics of the Use of Animals of the State University of Ponta Grossa (Protocol: 29/2016).

Sequences of the Gypsy obtained by PCR of the genomic DNA of *Apareiodon* sp. were labeled with digoxigenin-11-dUTP using the Dig-Nick Translation Mix Kit (Roche Applied Science). This Gypsy probe was used for FISH experiments with chromosome preparations of *A. piracicabae*, *Apareiodon* sp., *P. hilarii* and *A. affinis*. Signal detection was performed using the anti-digoxigenin rhodamine antibody, fab fragments (Roche Applied Science). The chromosomes were counterstained with 0.2 µg/ml of 4', 6-diamidino-2-phenylindole (DAPI) in Vectashield assembly medium (Vector, Burlingame, CA) and analyzed under light field microscopy and Zeiss Axio Lab 1 epifluorescence coupled with Zen Lite software on Zeiss AxioCam ICM1 camera, 1.4 megapixel resolution (Carl Zeiss®). The karyotypes were organized using Adobe Photoshop software, version CS6 and classified in metacentric (m), submetacentric (sm), subtelocentric (st) and acrocentric (a) (Levan et al. 1964).

### 3 RESULTS

#### 3.1 GENOME ASSEMBLY, REPETITIVE FRACTION IDENTIFICATION AND REPEAT LANDSCAPE ANALYSIS

In a previous study, Schemberger et al. (submitted for publication) showed that approximately 36% of the genome of *Apareiodon* sp. corresponds to the repetitive fraction, with the RTEs representing 27% of the total. Among these, the L1, L2, Rex-Babar and Gypsy were the most frequent and, despite the apparent molecular degeneration, Gypsy, Pao and ERV were the elements that presented the longest fragments (Schemberger et al. submitted for publication). In this study, we focused and selected sequences with similarity to Gypsy for molecular characterization.

The landscape analysis demonstrates Kimura nucleotide substitutions level ranging from 0 – 50% (Fig. 1). The major Gypsy copies number in male and female analysis demonstrated a nucleotide substitution level between 20 – 42% (Fig. 1).

Comparatively, the *Apareiodon* sp. female (ZW chromosomes) did not show a significantly differences when compared to male (ZZ chromosomes) in the landscape Kimura substitutions level analysis (Fig. 1).

### 3.2 SIMILARITY ANALYSIS

The analysis of RepeatMasker showed that the genome of *Apareiodon* sp. (genome available in <http://sacibase.ibb.unesp.br/jbrowse/JBrowse-1.12.1/index.html?data=data-apareiodon-sp>) has 36,732 contigs that presented sequences with identity to the Gypsy. These ranged from 10 to 5,611 bp in size (total length masked in the genome = 8,283,497 bp). All sequences showed a minimum similarity of 65% with the Gypsy of *Danio rerio*.

Sequences obtained from sequencing long-length PCR products returned six DNA fragments around 950 bp each, which were arranged in a consensus sequence of 5,325 bp. Sequence analysis showed 66% average similarity with the Gypsy of *Danio rerio* and *Gasterosteus aculeatus* and a similarity of 99.3% with the sequence present in the *Apareiodon* sp. contig 2. Analyzes of the individual sequencing obtained from the genome of *Apareiodon* sp. (sets of primers P1 to P6) also returned sequences with a similarity of 91.1% to the contig sequence 2.

The results in Plotcon demonstrated the Gypsy regions (Gag, Pol, LTR-5', LTR-3') low nucleotide similarity, in addition, the PNMA (Gag-Ty3/Gypsy-derived domain) showed medium similarity among its sequences (Supplementary Fig. 1A, 1B and 1C).

### 3.3 GYPSY SEQUENCES ANALYSIS

The corresponding *Gag* and *Pol* regions were detected in sequences of size  $\geq$  1,900 bp present in 145 contigs with Gypsy sequences (NCBI -Accession numbers: MK431240 - MK431376). After identity filtering, sequences from 24 contigs demonstrated structures not compatible with Gypsy and were removed from the analysis. Protein domains for the *Gag* region were found in 15 sequences out of the 121 analyzed, and at least one of the *Pol* domains was detected in 120 sequences (Fig. 2, Supplementary Table 4). Only two sequences (8 and 9) presented all domains

of the Gypsy (Fig. 2, Supplementary Table 4). In addition, *Gag* and *Pol* presented intact and degenerated domains (Table 1, Supplementary Table 4).

For similarity analysis with Gypsy LTRs, five sequences showed only LTR-5' (8, 9, 26, 31 and 64), 10 sequences only to LTR-3' (1, 3, 4, 5, 13, 44, 55, 57, 67, 68) and two sequences (27 and 33) presented the 5' and 3' LTRs (Fig. 3). In this analysis, the regions SIRs (short inverted repeats) TG / CA, U3 (3' single), TATA box (TATAA), polyadenylation signal (AATAAA, ATATAA or AGTAAA) and U5 (5' single) were detected. The sequences PBS (primer binding site), PPT (polypurine tail) and TSD (target site duplication) was not detected (Fig. 3).

The sequences showed regions for *Gag* and *Pol* in a single ORF or in different ORFs (Fig. 4). In addition, a domain belonging to a gene family derived from Gag-Ty3/Gypsy - PNMA (Paraneoplastic Ma antigens) was also found in this study (Figs. 2 and 4, Supplementary Table 4). Similarity analysis for the amino acid sequences allowed to identify protein domains of SCAN and motif zf-CCHC (zinc finger-Cysteine, Cysteine, Histidine, Cysteine) for the *Gag* polyprotein region and, AP (aspartic proteinase), RT (reverse transcriptase), RH (RNase H), INT (integrase) and (CHR) chromodomain, and a zf-HHCC subdomain (zinc finger-Histidine, Histidine, Cysteine, Cysteine) subclass for *Pol* polyprotein (Fig. 2, Supplementary Table 4). Analysis of the *Gag* polyprotein showed that the zf-CCHC domain exhibits 15-17 amino acid residues with the conservation of the C<sub>2</sub>C<sub>5</sub>H<sub>10</sub>C<sub>15</sub> positions in 9 sequences (Supplementary Table 4, Supplementary Fig. 2). For the SCAN domain, the sequences showed 86 to 89 amino acid residues. The presence of a partially leucine-rich nature was detected in the intact sequences of this domain, this amino acid being found in 9 positions (29, 45, 48, 50, 51, 66, 74, 80 and 86) (Supplementary Table 4, Supplementary Fig. 3). In the *Pol* polyprotein, the conserved region V / L<sub>14</sub>D<sub>15</sub>T<sub>16</sub>G<sub>17</sub>S<sub>18</sub> and L<sub>98</sub>L/I<sub>99</sub>G<sub>100</sub> (Supplementary Table 4 Supplementary Fig. 4) was found in the domain for AP, with 91 to 102 amino acid residues. In RT, intact domains were found with sequences of 155-163 amino acids and the presence of 7 conserved regions. Within these regions the 17 conserved amino acid positions could be detected, but 8 had identical residues in all sequences (positions 3, 15, 50, 85, 55, 115, 116, 149 and 163) and 9 showed variations in the sequences (22, 54, 55, 83, 84, 90, 113, 142 and 162). The positions conserved with chemically similar residues in 16 positions (1, 14, 17, 47, 58, 82, 93, 97, 110, 116, 117, 118, 119, 143, 145, 161 and 165) were also identified (Supplementary Table 4, Supplementary Fig. 5).

The intact sequences for the RH domain had 100-126 amino acids and conserved catalytic residues composed of three aspartic acids and one glutamic acid (D<sub>6</sub>E<sub>51</sub>D<sub>79</sub>D<sub>127</sub>) (Supplementary Table 4, Supplementary Fig. 6). The domain (INT) was determined to be intact in sequences ranging from 107-114 amino acids and preservation of the catalytic triad DDE (aspartate/aspartate/glutamate) at positions 10, 70 and 105, respectively (Supplementary Table 4, Supplementary Fig. 7). On the other hand, the CHR domain showed sequences ranging from 38-50 amino acid residues and 26 positions with identical residues and 4 positions with chemically similar residues (1, 6, 37 and 46) (Supplementary Table 4, Supplementary Fig. 8). A zf-HHCC subdomain associated with the (INT) domain was detected in the Pol region with 35-57 amino acid residues and, in seven sequences, the H<sub>14</sub>H<sub>21</sub>C<sub>51</sub>C<sub>54</sub> positions were determined (Supplementary Table 4, Supplementary Fig. 9). The intact PNMA was found in three sequences with similarity of 69.4% (Supplementary Table 4, Supplementary Fig. 10).

### 3.4 GENE EXPRESSION ANALYSIS

The RT and INT were not transcribed in male and female gonads of *Apareiodon* sp. On contrary, PNMA show an amplification band above 100 bp in testis of *Apareiodon* sp. in electrophoresis gel analysis (Supplementary Fig. 11).

### 3.5 IN SITU LOCATION

The *in situ* location revealed the presence of Gypsy dispersed along the chromosomes in the four species studied (*A. piracicabae*, *Apareiodon* sp., *Parodon hilarii* and *A. affinis*) without differential accumulations in the ZW sex chromosomes of *Apareiodon* sp. and *P. hilarii* and W<sub>1</sub>W<sub>2</sub> sex chromosomes of *A. affinis* (Fig. 5).

## 4 DISCUSSION

In fish genomes, the numbers and proportions of TEs are extremely variable, and in this group the largest number of superfamilies among vertebrates is visualized

(Duvernell et al. 2004; Kojima and Jurka 2011). Actinopterygians, in particular, have their genomes enriched with LTR-RTEs, being Gypsy considered one of the most abundant (Volf et al. 2001; Shen and Steiner 2004; Fischer et al. 2004; Kambol and Abtholuddin 2008; Gao et al. 2016). *Apareiodon* sp. has a relatively small genome (around 1.2 Gb) and the repetitive fraction annotation data demonstrated the RTEs are less frequent than DNA transposons (Schemberger et al. submitted for publication).

The repeat landscape showed one ancient wave of Gypsy insertion into the *Apareiodon* genome. It is due to a high level of degenerated Gypsy copies, probably defectives. In the cichlid *Astatotilapia latifasciata* the LTRs Gypsy also entered the genome in its early evolutionary stages and not was corelated to B chromosome transposable elements invasion (Coan and Martins 2018). Similarly, the landscape not evidenced important Gypsy invasion waves between male and female genome, without correlation in the sex chromosome differentiation in *Apareiodon* sp. The plotcon and landscape data demonstrate that the Gypsy element have numerous copies deteriorated in the *Apareiodon* genome. The invasion phase lead to copy number increase, after this time, the TE accumulate mutations over the generations causing inactivation and finally, its elimination (Kidwell and Lisch 2001; Fernández-Medina et al. 2012). In another pathway deteriorated TEs can generate co-opted copies to another genomic function (Feschotte 2008). The high level of the molecular deterioration suggests the non-autonomous, neutral or co-opted Gypsy in the *Apareiodon* genome.

The RTE-LTRs have their LTRs sequences flanking the coding region, which contain promoters and consensus sequences necessary for retrotransposition (Vernhetes et al. 1997; Sabot and Schulman 2006; Li et al. 2019). When these RTEs are active, the LTRs retain high similarity. However, over the generations there is accumulation of mutations, with consequent reduction of similarity (Gao et al. 2003).

In this study it was possible to detect sequences with similarity to LTR-5' and LTR-3' of Gypsy. These consensus regions of LTR-5' and LTR-3', when compared to each other, showed similarity, however the sequences between them presented innumerable single nucleotide polymorphisms (SNPs) and Indels (insertion and deletion). Sperber et al. (2007) reported that the isolated analysis of the candidate consensus sequences for LTRs can generate an unsatisfactory result with false positives. However, the absence of intact LTR-5' and LTR-3' regions in Parodontidae Gypsy may be a consequence of the molecular deterioration of this RTE in the

genome. Based on the level of sequence deterioration and non-identification of the promoter and the transcription termination site located in the LTRs, it is also possible to infer that the Gypsy sequences of Parodontidae are not able of mobilization, as proposed by Gao et al. (2003) in a study of the characterization of RTE-LTR from other organisms.

In addition to the LTR-5' and LTR-3' regions, the regions corresponding to Gypsy ORF from *Apareiodon* sp. also demonstrated numerous SNPs and Indels. Sequences corresponding to *Gag* and *Pol* of Gypsy of *Apareiodon* sp. were detected by organizing a single ORF or in up to 4 different ORFs. It is worth noting that in most RTEs, *Pol* and *Gag* do not have independent translation initiation, being expressed as part of a *Gag-Pol* polyprotein (Gao et al. 2003; Wicker et al. 2007; Benachenhou et al. 2013). This mechanism is important for the packaging of *Pol* products within the particle. However, for assembly of the particles, more copies of *Gag* relative to *Pol* are required and therefore the level of expression of *Pol* occurs through deviation from the standard coding mechanism, at different reading phases +1 and -1 (Hull and Covey 1995; Shehu-Xhilaga et al. 2001; Telenti et al. 2002; Gao et al. 2003). All analyzes have shown that the molecular deterioration of these sequences is compatible with a long generation time without the mobilization of this RTE in the genome.

In the Gypsy sequences of *Apareiodon* sp. the regions of similarity to *Gag* and *Pol* were identified and characterized, although in many cases, with truncated domains. Protein domains of *Gag* are essential for the formation of mRNA encapsulation particles in retroviruses, with distinct roles during viral assembly (Morita and Sundquist 2004, Cao et al. 2016). However, in studies of characterization of Gypsy, the *Gag* region is characterized by the presence of one or more zinc fingers, based mainly on the conserved domain CCHC, which binds to RNA of the RTE, similar to that found in retroviral particles (Cao et al. 2016). In this study, the Gypsy *Gag* region of *Apareiodon* sp. demonstrated a conserved zf-CCHC domain, exhibiting the least divergence between amino acid sequences when compared to the other protein domains.

Furthermore, the domain known as SCAN, which belongs to the retrovirus capsid dimerization superfamily (Edelstein and Collins 2005) was detected in the *Gag* region of the analyzed genome. This domain is a subfamily associated with zf-HHCC modules and involved in protein-protein interactions, with possible participation in the assembly of a family of regulatory proteins (Williams et al. 1999). In our analysis, this



domain was detected associated with the zf-CCHC domain or isolated, always located upstream of the zf-CCHC domain region.

In the Gypsy, the order of the *PoI* domains presents the sequence AP-RT-RH-INT, which helps in differentiation in relation to the other LTRs (Wicker et al. 2007). The four domains that make up *PoI* have different functions in the retrotransposition process. The AP domain of Gypsy of *Apareiodon* sp. presented the two conserved V/L DTG and LL/IG motifs (Doolittle et al. 1989) for some sequences, in addition to a high divergence in amino acid residues in the other regions. These proteolytic enzymes play a key role in the maturation process of other peptides (Rawlings et al. 2008).

The RT domain was the most observed in the Gypsy sequences of *Apareiodon* sp., where it was possible to determine seven regions of conserved amino acids for the function of RT (Saigo et al. 1984, Xiong and Eickbush 1990) in ~ 33% of the sequences. Already for the RH domain, in Gypsy sequences of *Apareiodon* sp. which were shown to be intact, showed the conserved DEDD domain, essential for their activity (Johnson et al. 1986; Doolittle et al. 1989). For the INT domain, approximately half of the Gypsy sequences of *Apareiodon* sp. demonstrated the DDE motif, which is essential in the activity of the enzyme in catalyzing the insertion of the new copy of the RTE-LTR in the genome (Chiu and Davies 2004; Nowotny 2009). Despite some few copies show RT and IN proteic domains, no transcription was detected for these Gypsy segments in gonads, corroborating to proposal of deteriorated copies and lack of mobilization this TE in the genome.

The subdomain zf-HHCC, detected upstream of the integrase domain, participates in the process of cleavage and integration of the new RTE-LTR into the genome, but its exact function remains unknown. In our study, we detected some mutations in this motif that may be reducing zinc binding and damaging the activity of this subdomain (Bushman et al. 1993; Lodi et al. 1995; Cai et al. 1997; Eijkelenboom et al. 1999). Elements belonging to the Metaviridae strain of Gypsy retrotransposons may further exhibit the CHR domain in their structure. Chromo is involved in the remodeling of chromatin and the regulation of gene expression in eukaryotes (Koonin et al. 1995; Cavalli and Paro 1998). In the present study, this domain was found for some Gypsy sequences of *Apareiodon* sp. at the C-terminal end of the integrase, which is characteristic of this domain by participating in conjunction with the integrase of the Gypsy integration process in the genome (Koonin et al. 1995; Cavalli and Paro 1998).



Thus, although the evidence analyzed demonstrates the likely inactivity of *Gag* and *Pol* de Gypsy de *Apareiodon* sp., some protein domains still retain their structure and thus remain neutral in the genome or otherwise could be co-opted for other functions, as already proposed for proteins derived from TEs (Feschotte 2008). An example of molecular co-option of part of TE is the PNMA domain, which is derived from *Gag* and found in the mammalian X chromosome (Voff 2009). Some of these proteins retained the ancestral *Gag* zinc finger motif, but others have lost that domain (Voff 2006). Gypsy sequences of *Apareiodon* sp. which show similarity to the PNMA domain were detected in our study. In addition, PNMA was transcribed in the testis of *Apareiodon*.

In humans, the onconeural proteins PNMA1 and PNMA2 have been identified as targets of immunity associated with paraneoplastic neurologic syndromes (Dalmau et al. 1999). Later, Rosenfeld et al. (2001) identified three types of the PNMA proteins (Ma1, Ma2, Ma3). These proteins in humans are selectively expressed in neurons and tumors, Ma1 is also expressed in germ cells of the testis and, Ma3 ubiquitously expressed in brain, testis, and several systemic tissues (Rosenfeld et al. 2001). In other study, Iwasaki et al. (2003) identified two novel candidate PNMA genes in marsupials, which ones were described to exhibit the highest homology to a Gypsy12 (*Gag* protein) of the fish *Danio rerio*. Since Gypsy sequences have high nucleotide substitution level and promoter region are not detected in LTR analysis, the transcription visualization of PNMA suggest other genomic functions to Gypsy deteriorated copies in *Apareiodon* sp., maybe in other fish too. In this case, the PNMA origin or, mutations stages previously to PNMA emergence occurred in ancient evolutionary lineages before mammals diversification.

The *in situ* location of the Gypsy was dispersed along the chromosomes of the *A. piracicabae* (without sex chromosomes) and *Apareiodon* sp., *Parodon hilarii* and *Apareiodon affinis* (species that have heteromorphic sex chromosomes ZZ/ZW or ZZ/ZW<sub>1</sub>W<sub>2</sub>, for a review see Schemberger et al. 2011) and we did not detected differential accumulations of Gypsy between the Z and W chromosomes. These chromosomal evidences together to ancient genomic invasion of Gypsy in landscape analysis comprove the inactivity of this element during the differentiation of the sex chromosomes in Parodontidae. These results also coincide with the *in situ* location of Gypsy in other fishes, where they occur scattered throughout the length of the

chromosomes, with some points of accumulation in the nucleolar, centromeric and telomeric organizing regions (Fischer et al. 2005; Cao et al. 2016).

TEs are recognized to promote genomic reshuffling and provide evolutionary novelties in co-opted functions. The results obtained with the characterization of Gypsy in conjunction with the *in situ* location data allow us to infer that Gypsy entered the Parodontidae genome in its early evolutionary stages, before sex chromosomes differentiation. Yet, deteriorated Gypsy copies may have originated raw DNA material to PNMA emergence before mammals diversification. The high molecular deterioration demonstrate the lack of genomic Gypsy mobilization in *Apareiodon*.

## ACKNOWLEDGEMENTS

The authors are grateful to ICMBio (Instituto Chico Mendes de Conservação da Biodiversidade) (license number 10538-1 to collect specimens). This study was supported by Fundação Araucária (Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico do Estado do Paraná), FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, process: 2015/16661-1), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, - Finance Code 001) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

## CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

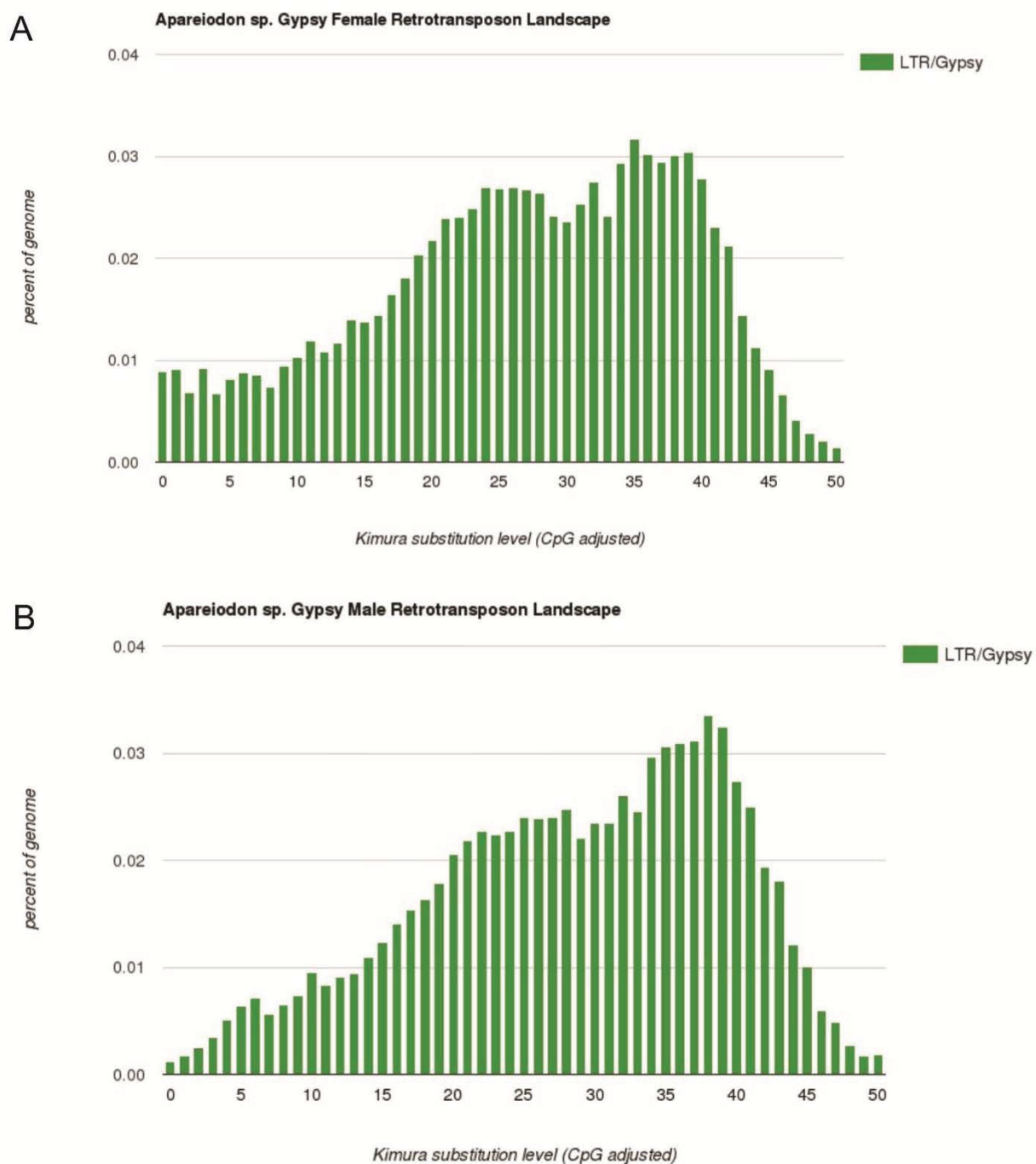
## REFERENCES

As referências utilizadas neste artigo se encontram ao final da tese.

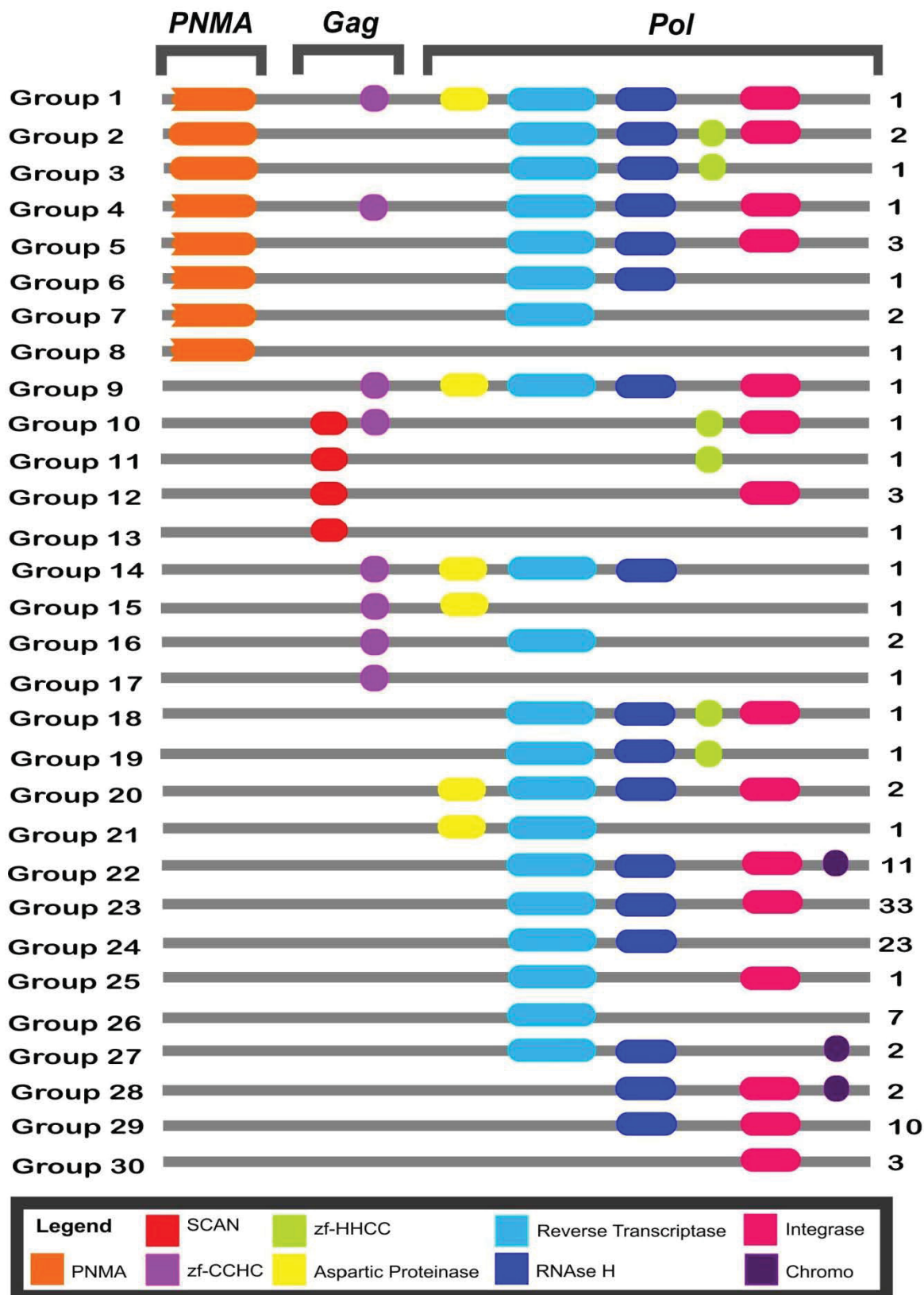
**Table 1** Whole and degenerate protein domains of Gypsy detected in *Apareiodon* sp. Data based on 121 contigs with size  $\geq 1900$  bp.

	Protein domains	Intact	Degenerate	Total
GAG	Scan	5	1	6
	Zf – CCHC	9	0	9
POL	Aspartic proteinase	6	1	7
	Reverse transcriptase	33	64	97
	RNAse H	53	43	96
	Zf – HHCC	7	0	7
	Integrase	38	37	75
	Chromo	15	0	15
	PNMA	3	9	12

**Fig. 1** Landscape analysis of the RTE-Gypsy in *Apareiodon* sp. genome. The graphs show the Gypsy divergence from its consensus (x-axis) in relation to the number of copies on the genome (y-axis). Peaks represent insertion waves of Gypsy into the genome. In a) female and, b) male.

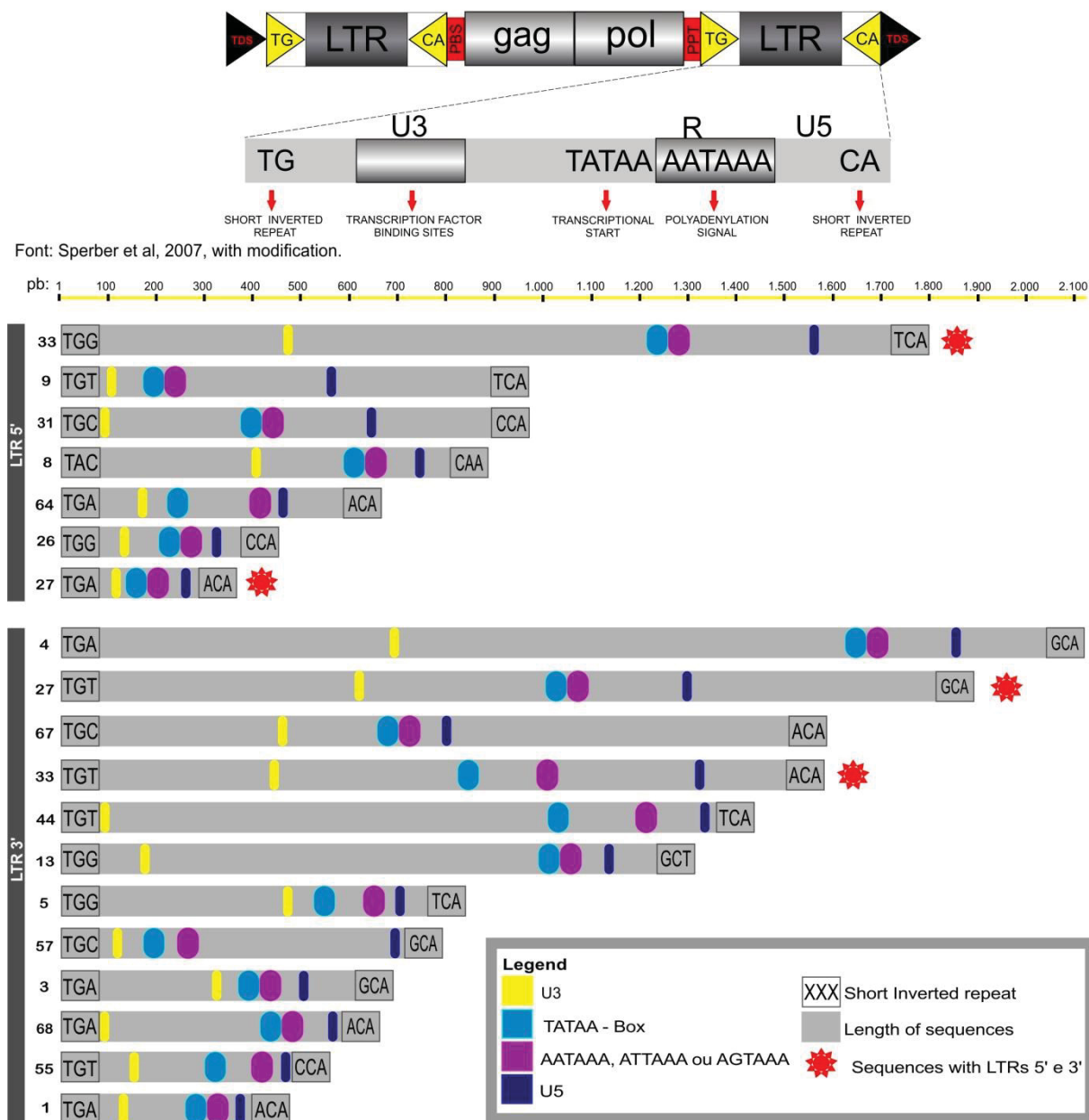


**Fig. 2** Schematic representation of the possible combinations of the protein domains of Gypsy found in *Apareiodon* sp., in the 121 contigs analyzed. The numbering on the right indicates the number of times this setting was detected. Distances between domains, size, integrity, and location do not match the original sequences.



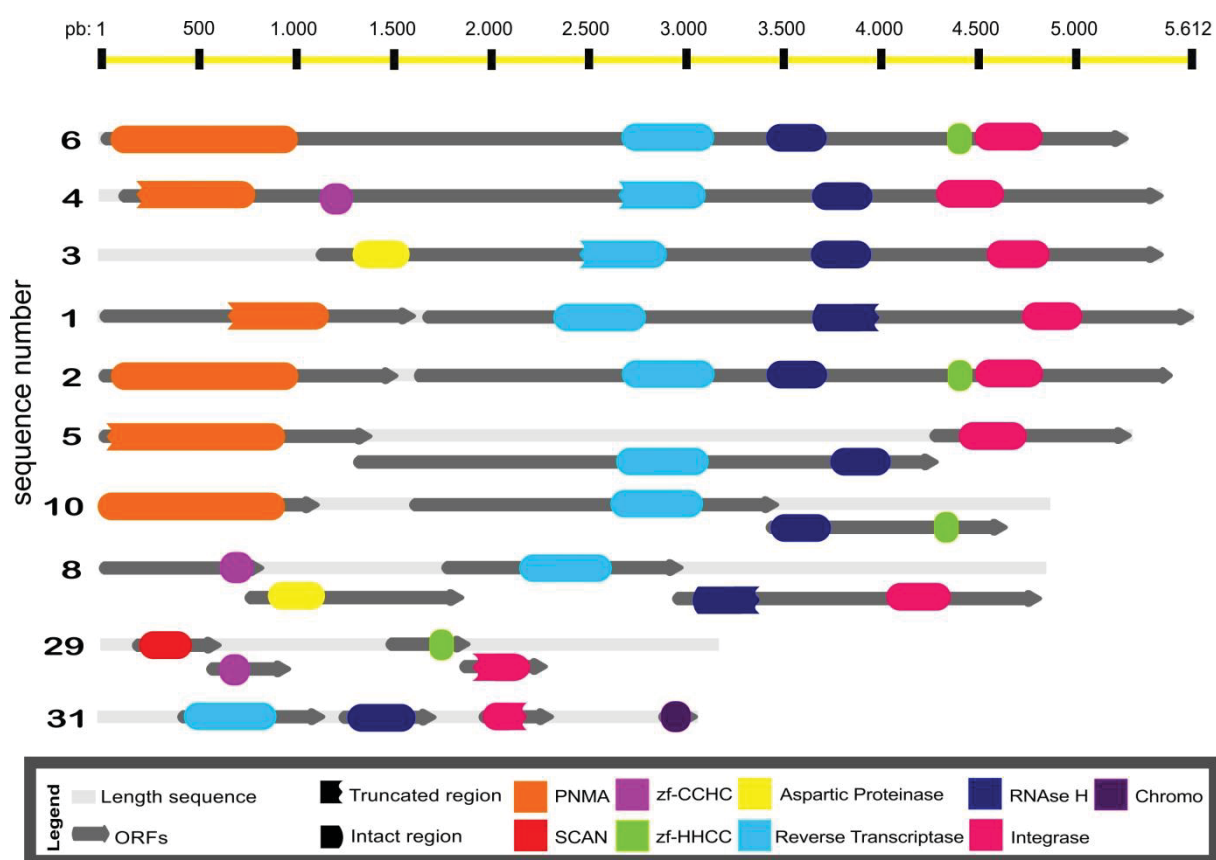
**Fig. 3** Schematic representation of Gypsy LTRs, detected in Retrorector (Sperber et al. 2007), showing the main conserved regions.

Schematic drawing of Gypsy structure

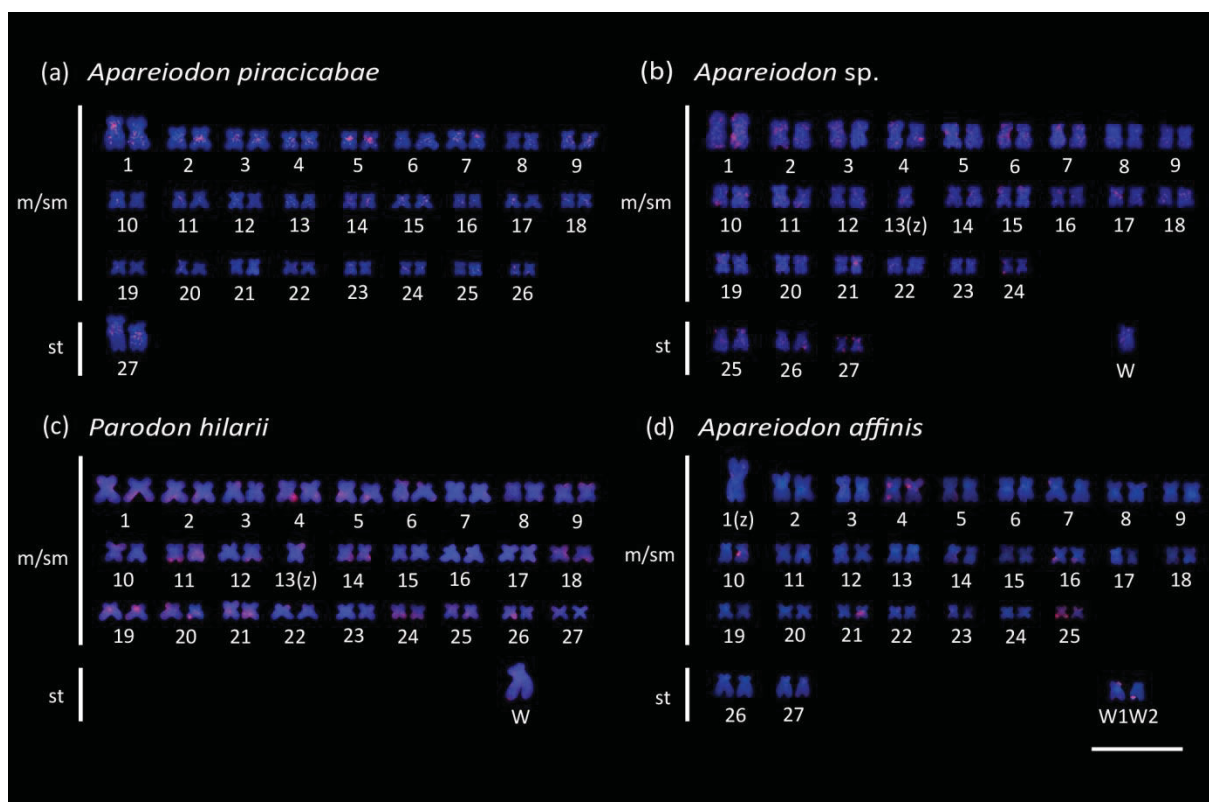




**Fig. 4** Representation of possible types of ORFs detected in 121 contigs analyzed for Gypsy in *Apareiodon* sp. Sequences with 1, 2, 3 and 4 ORFs were founded.



**Fig. 5** *In situ* location of the Gypsy RTE and karyotypes arrange in female individuals of Parodontidae species: *A. piracicabae*, *Apareiodon* sp. (a), *Parodon hilarii* (b) and; *Apareiodon affinis* (c). FISH using Gypsy probe (red) obtained from *Apareiodon* sp. (set of primers 3).



**Supplementary Table 1** Primer sets constructed for the amplification and characterization of the RTE-LTR sequences Gypsy, using the software Primer3Plus. Expected fragment size ~ 950 bp for each set.

IDENTIFICATION		SEQUENCES
<b>Set 1</b>	Gypsy_P1_Forward	5' GCTAGCAAGTGGTCTCTATGAA 3'
	Gypsy_P1_Reverse	5' AGGCACAGATAGGGCTATAGAA 3'
<b>Set 2</b>	Gypsy_P2_Forward	5' CCCCACAGCTCTTTATGTAGT 3'
	Gypsy_P2_Reverse	5' TGGAGAGAGGTGGACTGAA 3'
<b>Set 3</b>	Gypsy_P3_Forward	5' GGTCAGCAAAAGCCAGTAG 3'
	Gypsy_P3_Reverse	5' CACATCAGGCTCAAGGACT 3'
<b>Set 4</b>	Gypsy_P4_Forward	5' CAGCTAGCAGCTCCTGAA 3'
	Gypsy_P4_Reverse	5' GGATAGCGGGTCTCAAGT 3'
<b>Set 5</b>	Gypsy_P5_Forward	5' TGAATGAACCCTTTATTGTCA 3'
	Gypsy_P5_Reverse	5' CTAAGGAGGTTGGAGCAAG 5'
<b>Set 6</b>	Gypsy_P6_Forward	5' GTTACTGCACCTCTCAAAAGC 3'
	Gypsy_P6_Reverse	5' GGGGAGGGCCTAAGAGTA 5'

**Supplementary Table 2** Primer sets constructed for the expression analysis of the RTE-LTR sequences Gypsy, using the software Primer3Plus.

IDENTIFICATION		SEQUENCE	FRAGMENT SIZE
<b>Integrase</b>	Forward	5' GGCCCTCATTTTGAAAGCC 3'	91pb
	Reverse	5' ACCCTGAGGATGATAAGCTGTG 3'	
<b>Reverse Transcriptase</b>	Forward	5' GCTCACTAATGCCCCAGCTAC 3'	175pb
	Reverse	5' CTTTAGTCCATCCTCCTGCAGTC 3'	
<b>PNMA</b>	Forward	5' AGCAGGGGAAAGAAGGCTCT 3'	129pb
	Reverse	5' ATTCCTAAACCAAGTGGTGAAGTTG 3'	

**Supplementary Table 3** Parodontidae species name, sampling locality, distribution in river basins and number of species studied.

Species	River, state	Hydrographic basin	Number of ♂	Number of ♀	GPS
<i>A. piracicabae</i>	Mogi-Guaçu River - SP	Alto Paraná	8	6	-21°55'38.6" and -47°21'58.4"
<i>Apareiodon</i> sp.	Verde River - PR	Alto Paraná	14	12	-25°04'35" and -50°04'03"
<i>P. hilarii</i>	São Francisco River - SP	São Francisco	12	8	-17°21'17" and -44°57'15"
<i>A. affinis</i>	Passa Cinco River - SP	Alto Paraná	10	7	-22°25'26" and -47°41'56"

Supplementary Table 4 Description of the protein domains found in the 145 contigs analyzed from *Apareiodon* sp.

Contig number	Size in pb	ORFs Number	GAG			POL					
			PNMA	SCAN	Zf-CCHC	Aspartic Proteinase	Reverse transcriptase	RNAse H	Zf-HHCC	Integrase	Chromo
1	5612	2	489/163*	-	-	-	480/160	237/79*	-	336/112	-
2	5500	2	963/321	-	-	-	477/159	300/100	168/56	333/111	-
3	5458	1	-	-	-	282/94	474/158	369/123	-	342/114	-
4	5348	1	651/217*	-	48/16	-	372/124*	369/123	-	342/114	-
5	5306	3	909/303*	-	-	-	474/158	357/119	-	336/112	-
6	5286	1	963/321	-	-	-	477/159	300/100	168/56	333/111	-
7	4976	1	483/161*	-	-	-	480/160	363/121	-	324/108	-
8	4826	4	-	-	51/17	306/102	483/161	369/113*	-	339/113	-
9	4762	4	420/140*	-	48/16	300/100	483/161	363/121*	-	321/107*	-
10	4724	3	969/323	-	-	-	480/160	315/105	168/56	-	-
11	4486	1	-	-	-	-	369/123*	366/122	-	-	-
12	4165	2	-	-	-	-	261/87*	363/121*	-	336/112	-
13	3993	1	-	-	-	-	480/160	360/120	-	336/112	-
15	3971	1	-	-	-	-	480/160	357/119	-	333/111	-
16	3905	2	501/167*	-	-	-	480/160	357/119	-	-	-
17	3711	2	-	267/89	-	-	-	-	-	294/98*	-
18	3594	1	-	-	-	-	408/136*	366/122	-	342/114	-
19	3487	2	-	-	-	-	390/130*	294/98*	-	339/113	-
20	3477	3	-	-	-	-	474/158	372/124	-	333/111	-
21	3412	2	-	-	-	-	153/51*	369/123	111/37	288/96*	-
22	3410	3	-	-	-	-	381/127*	294/98*	-	321/107*	-
23	3369	3	-	-	48/16	303/101	483/161	336/112*	-	-	-
24	3366	1	-	-	-	-	468/156	-	-	-	-
25	3347	2	954/318*	-	-	-	348/116*	-	-	-	-
26	3289	1	-	-	-	-	482/161	369/123	-	342/114	-
27	3215	1	-	-	-	-	480/160	363/121*	-	339/113	-
28	3182	3	-	-	-	-	480/160	315/105	-	330/110*	-
29	3181	4	-	258/86	48/16	-	-	-	168/56	285/95*	-
30	3154	1	-	-	-	-	279/93*	126/42*	-	-	-
31	3140	4	-	-	-	-	483/161	366/122	-	276/92*	150/50

Continuation Supplementary Table 4

32	3097	2	-	-	-	-	-	366/122*	345/115	-	258/86*	-
33	3095	2	-	-	-	-	-	357/119*	354/118	-	336/112	117/39
34	3075	2	-	-	-	-	-	153/51*	153/51 e	-	339/113	-
35	3060	2	-	-	-	-	-	483/161	342/114*	-	-	144/48
36	3050	3	-	-	-	-	-	483/161	354/118	-	339/113	-
37	3040	2	-	-	-	-	-	480/160	-	-	-	-
39	3005	3	-	-	-	-	-	483/161	369/123	-	336/112	150/50
41	2989	4	-	-	-	-	-	465/155	363/121	-	339/113	144/48
42	2974	3	-	-	-	-	-	-	117/39 e	-	339/113	144/48
44	2963	4	-	-	-	-	-	300/100*	228/76 e	-	180/60*	147/49
45	2934	2	-	-	-	-	-	342/114*	240/80*	-	-	-
46	2879	1	-	-	-	-	-	-	-	-	252/84*	-
47	2875	1	-	-	-	-	-	378/126*	-	-	-	-
48	2872	3	-	-	-	-	-	231/77*	183/61 *e	-	-	-
49	2866	3	-	-	-	-	-	489/163	306/102	-	321/107*	-
50	2839	1	-	-	-	-	-	162/54*	348/116*	-	339/113	144/48
51	2834	2	-	-	-	-	-	399/133*	363/121	-	342/114	-
52	2831	1	-	-	-	-	-	366/122*	-	-	-	-
53	2821	3	-	-	-	-	-	249/83*	168/56*	-	339/113	141/47
54	2810	3	-	-	-	-	-	219/73*	99/33*	-	336/112	-
55	2808	4	-	-	-	-	-	183/61*	366/122	-	255/85*	-
56	2800	1	-	-	-	-	-	276/92*	366/122	-	285/95*	-
57	2767	4	-	-	-	-	-	363/121*	339/113*	-	300/100*	150/50
58	2758	3	-	-	-	-	-	240/80*	363/121	-	162/54*	-
59	2754	3	-	-	-	-	-	483/161	273/91*	-	339/113	-
60	2735	2	-	-	-	-	-	-	219/73*	-	234/78*	-
61	2723	2	-	-	-	-	-	303/101* e	-	-	-	-
62	2723	1	-	-	-	-	-	480/160	351/117	-	333/111	-
63	2692	3	-	-	-	-	-	216/72*	360/120	-	204/68*	-
64	2687	4	-	-	-	-	-	357/119*	135/45* e	-	198/66*	120/40
65	2684	2	-	-	-	-	-	165/55*	363/121	-	-	-
66	2663	2	-	-	-	-	-	450/150*	-	-	-	-
67	2653	3	-	-	-	-	-	483/161	369/123	-	201/67*	144/48
68	2644	3	-	-	-	-	-	357/119*	234/78* e	-	282/94*	-

Continuation Supplementary Table 4

69	2631	1	-	-	-	-	-	300/100*	351/117	-	-	-	-
70	2611	2	-	-	-	-	-	480/160	351/117	-	-	240/80*	-
71	2604	1	-	-	-	-	-	249/83*	360/120	-	-	-	-
73	2584	1	-	267/89	-	-	-	-	-	-	-	333/111	-
75	2579	1	-	-	-	-	-	477/159	369/123	-	-	-	-
76	2559	3	-	-	-	-	-	168/56*	360/120	-	-	186/62*	-
77	2553	2	-	-	-	-	-	-	264/88*	-	-	336/112	-
78	2533	3	-	-	-	-	-	216/72*	366/122	-	-	201/67*	-
80	2531	4	-	-	-	-	-	189/63*	267/89*	-	-	321/107*	144/48
81	2531	1	-	-	-	-	-	273/91*	-	-	-	-	-
82	2518	3	-	-	-	-	-	168/56*	243/81*	-	-	303/101*	-
83	2506	2	291/97*	-	-	-	-	309/103*	-	-	-	-	-
87	2485	3	-	-	-	-	-	216/72*	354/118	105/35	-	-	-
88	2480	1	-	-	-	-	-	267/89*	345/115*	-	-	267/89*	-
89	2433	2	-	-	-	-	-	-	351/117	-	-	237/79*	-
90	2421	2	-	-	-	-	-	237/79*	261/87*	-	-	-	-
91	2409	2	-	-	-	-	-	165/55*	225/75*	-	-	-	-
92	2375	3	-	-	-	-	-	258/86*	357/119	-	-	-	147/49
93	2375	2	-	-	-	-	-	390/130*	342/114	-	-	-	-
94	2370	1	-	-	-	-	-	375/125*	363/121	-	-	333/111	-
95	2370	2	-	-	-	-	-	357/119*	360/120*	-	-	-	-
96	2365	3	-	-	-	-	-	153/51*	243/81*	-	-	339/113	-
97	2358	3	-	-	-	-	-	414/138*	366/122	-	-	147/49*	-
98	2339	2	-	-	-	-	-	225/75*	189/63*	-	-	-	-
102	2331	2	-	-	-	-	-	-	228/76*	-	-	135/45*	-
103	2328	2	-	-	-	-	-	-	273/91*	-	-	237/79*	-
104	2277	1	-	-	-	-	-	-	-	-	-	282/94*	-
105	2260	1	-	-	-	-	-	480/160	363/121	-	-	-	-
107	2258	3	-	-	-	-	-	363/121*	117/39*e	-	-	144/48*	-
109	2257	2	-	-	-	-	-	195/65*	354/118	-	-	-	-
110	2249	2	-	-	-	-	-	342/114*	366/122*	-	-	-	-
112	2218	1	-	-	-	-	-	375/125*	372/124	-	-	-	-
113	2214	2	-	-	-	-	-	375/125*	333/101*	-	-	-	-
115	2168	2	-	-	-	-	-	234/78*	357/119	-	-	-	-

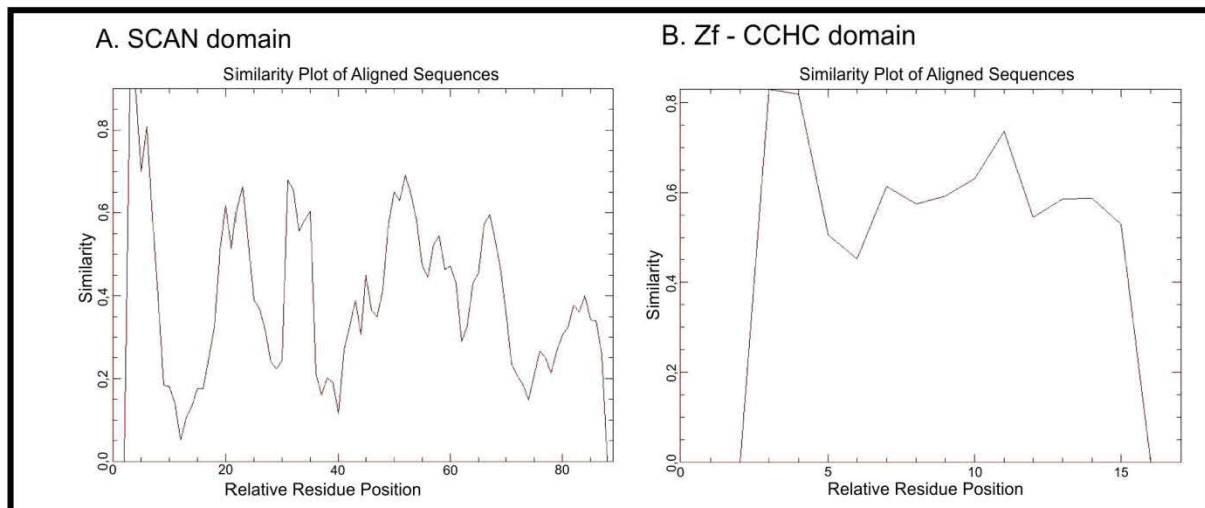


Continuation Supplementary Table 4

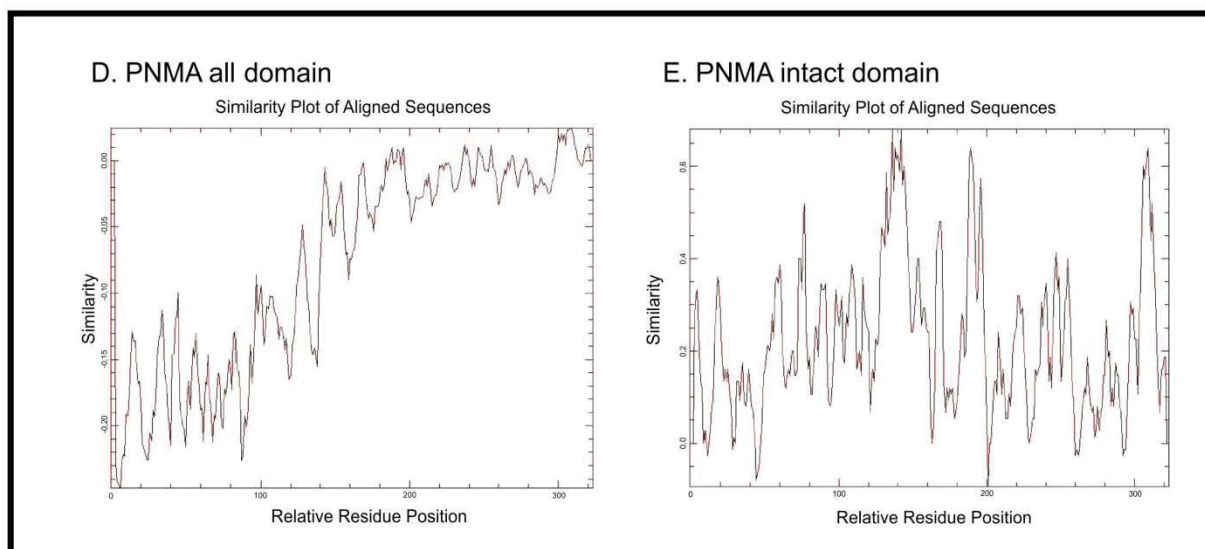
116	2163	3	-	-	-	165/55*	-	-	-	-	-	-
117	2150	3	-	-	-	354/118*	-	-	171/57* e	-	-	-
118	2146	2	-	-	-	-	-	-	360/120	-	336/112	-
119	2137	1	-	-	-	-	-	-	363/121	-	336/112	-
120	2134	3	-	-	-	135/45*	-	-	342/114*	-	336/112	-
121	2123	1	-	-	264/88	-	-	-	-	-	-	-
122	2118	2	-	-	-	-	-	-	366/122	-	135/45*	-
124	2107	2	-	-	-	267/89*	-	-	348/116	-	-	-
126	2096	2	-	-	-	-	-	-	183/61*	-	336/112	-
127	2078	1	-	-	-	273/91	-	-	-	-	-	-
128	2063	1	-	-	-	-	-	-	372/124	-	321/107	-
131	2054	2	-	-	159/53*	-	-	-	-	-	204/68*	-
132	2035	1	-	-	-	-	-	-	366/122	-	330/110*	96/32
134	1999	1	-	-	-	-	-	-	-	-	-	-
135	1993	3	-	-	-	207/69*	-	-	366/122*	-	-	-
136	1982	2	-	-	-	231 /77* e 177/	-	-	-	-	-	-
137	1971	3	-	-	-	306/102	-	-	-	-	-	-
139	1961	1	420/140*	-	-	363/121*	-	-	366/122	-	285/95*	-
141	1959	1	-	-	-	468/156	-	-	273/91*	-	-	-
142	1939	1	-	-	-	255/85*	-	-	120/40*	-	-	-
143	1905	1	-	-	264/88	-	-	-	-	171/57	-	-
144	1903	1	-	-	-	-	-	-	-	-	195/65*	-
145	1900	1	-	-	-	48/16	-	-	-	-	-	-

**Supplementary Fig. 1A** . Plotcon graphs demonstrating the similarity of the sequences for each GAG region domain (frame 1) and PNMA (frame 2) in *Apareiodon* sp. (in D, alignment using 12 PNMA of the *Apareiodon* sp. and; in E, alignment using integre PNMA of *Apareiodon* with PNMA of mammals).

Frame 1

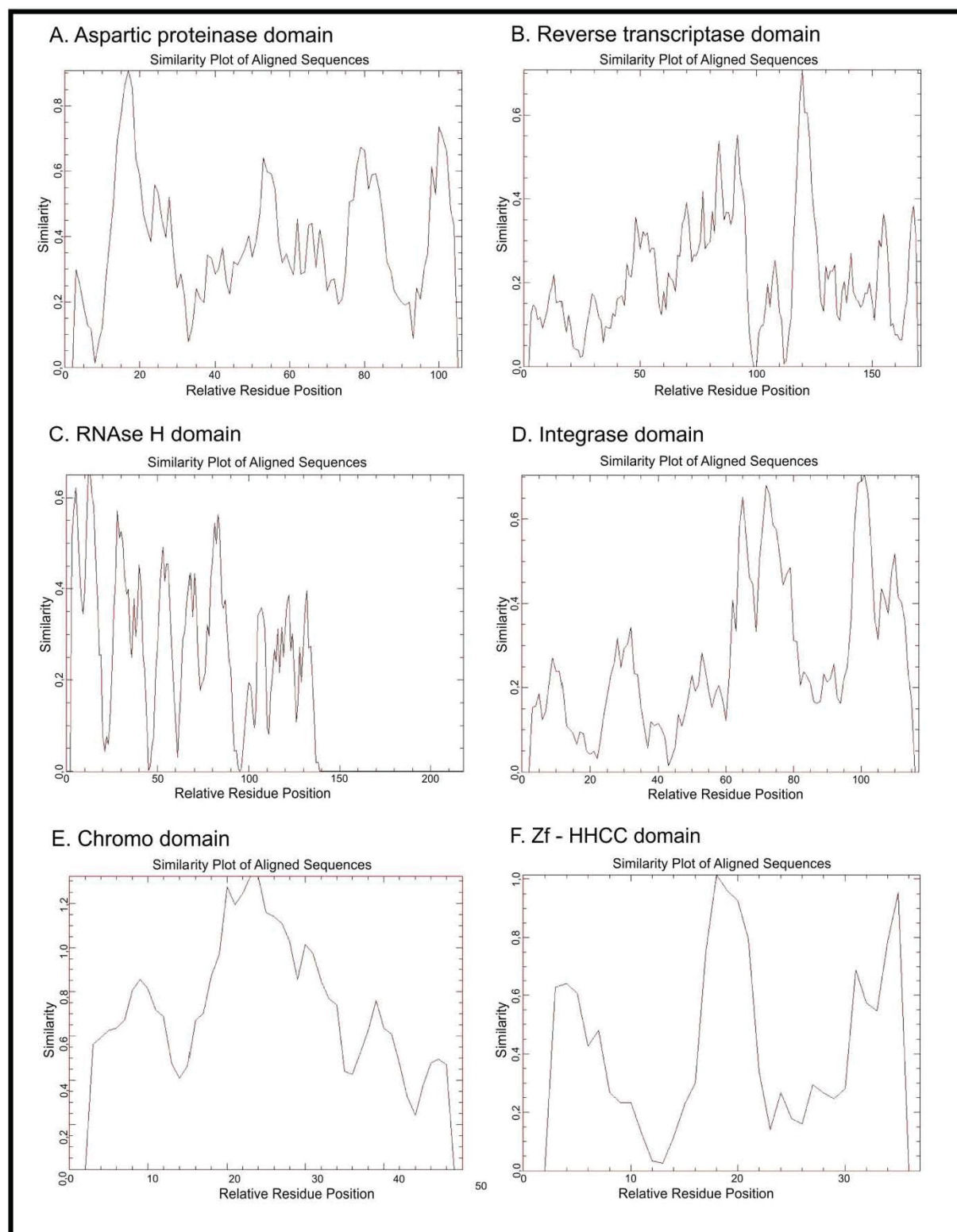


Frame 2

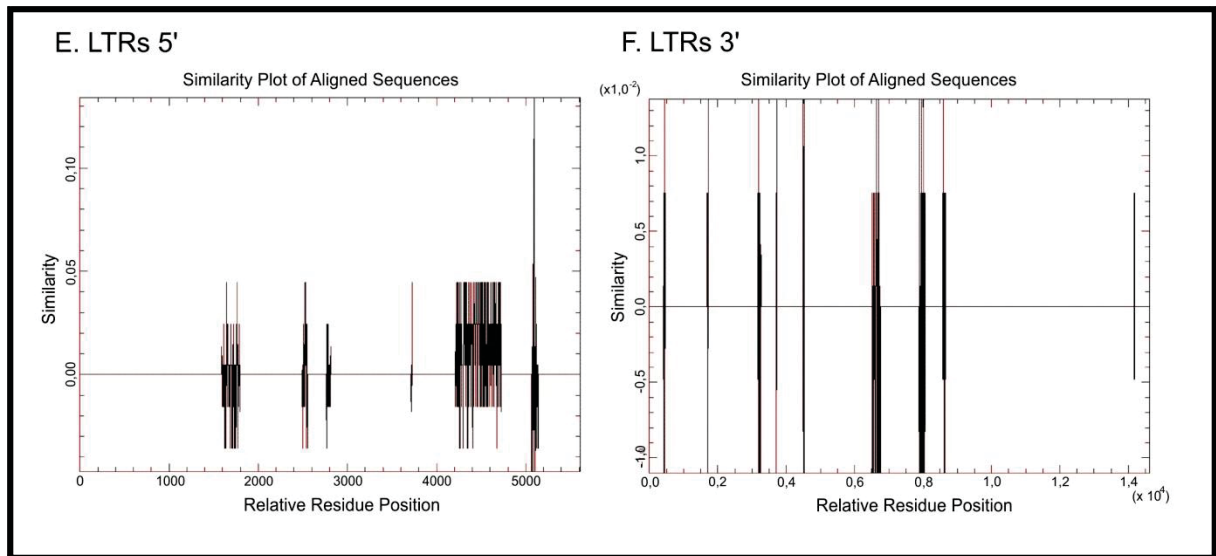


**Supplementary Fig. 1B.** Plotcon graphs demonstrating the similarity of the sequences for each protein domain of POL in *Apareiodon* sp.

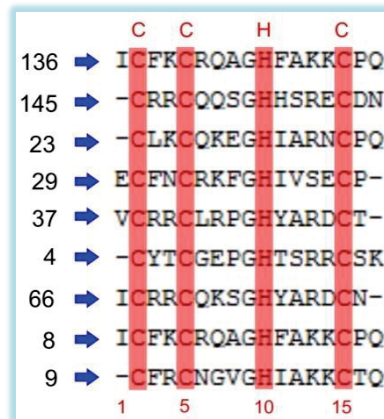
Frame 1



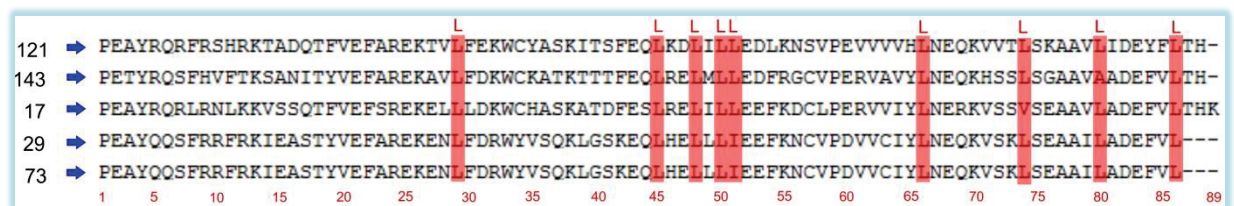
**Supplementary Fig. 1C.** Plotcon graphs demonstrating the similarity of the 5' and 3' LTR sequences in *Apareiodon* sp.



**Supplementary Fig. 2** Alignment MUSCLE of the intact Zf - CCHC sequences showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of *Apareiodon* sp.

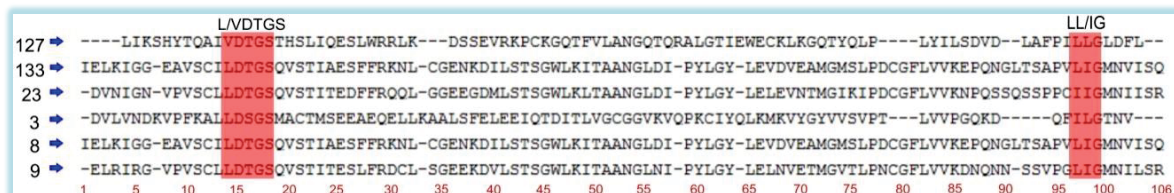


**Supplementary Fig. 3** MUSCLE alignment of the SCAN intact sequences evidencing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of *Apareiodon* sp.





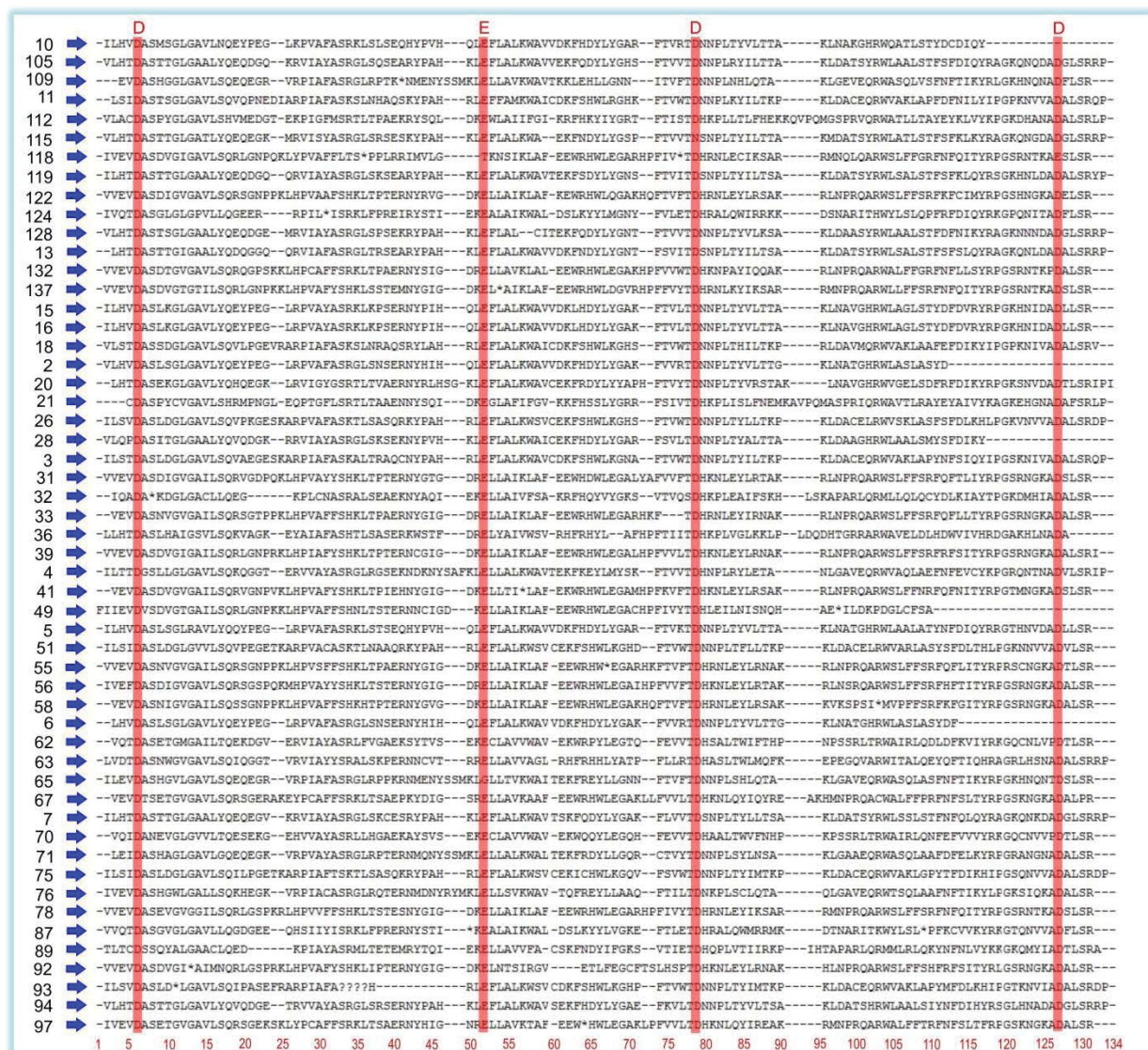
**Supplementary Fig. 4** Alignment MUSCLE of the intact sequences of Aspartic Proteinase showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of *Apareiodon* sp.



**Supplementary Fig. 5** Alignment MUSCLE of the Reverse transcriptase intact sequences showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of *Apareiodon* sp.

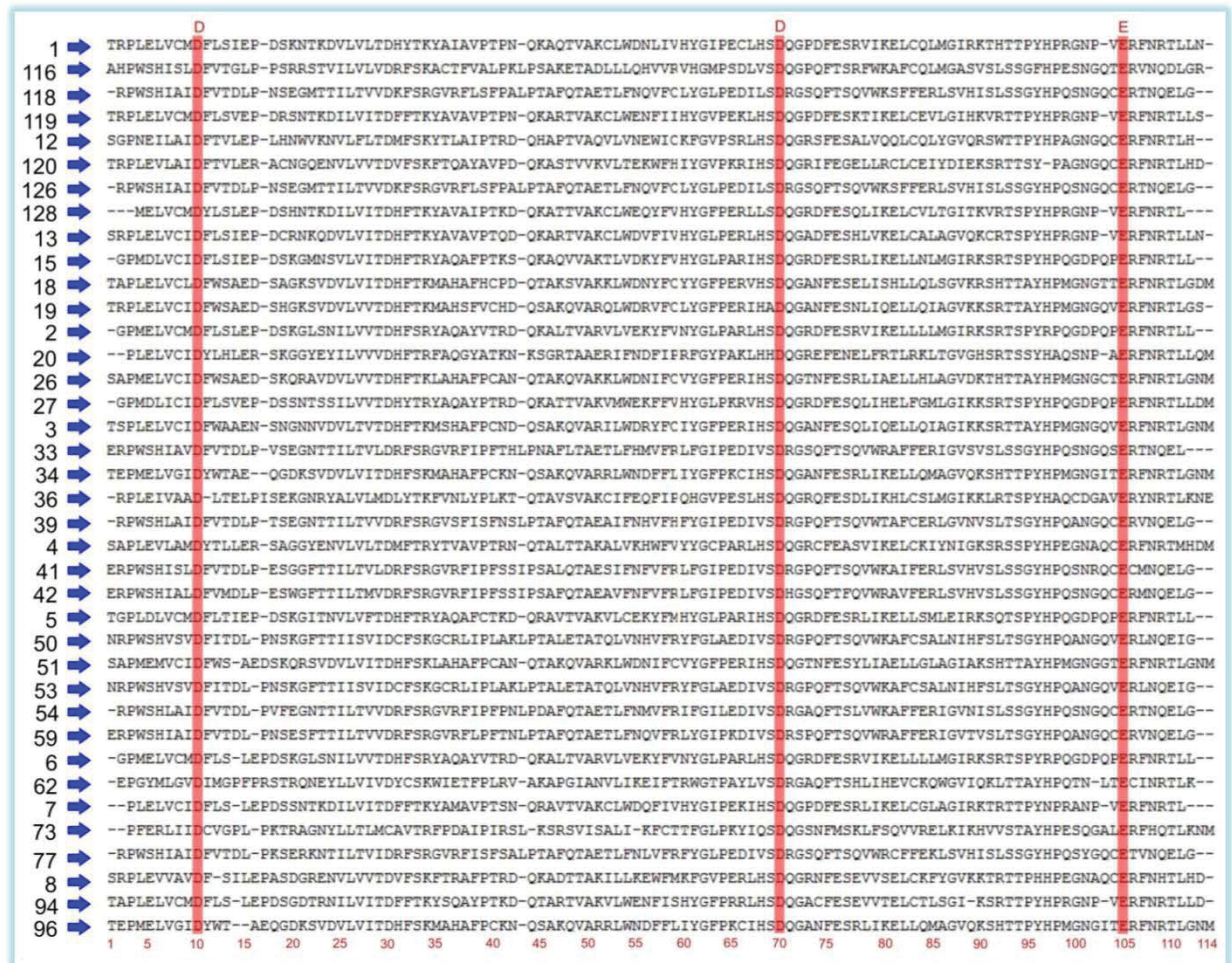


**Supplementary Fig. 6** Alignment MUSCLE of the RNase H intact sequences showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of *Apareiodon* sp.



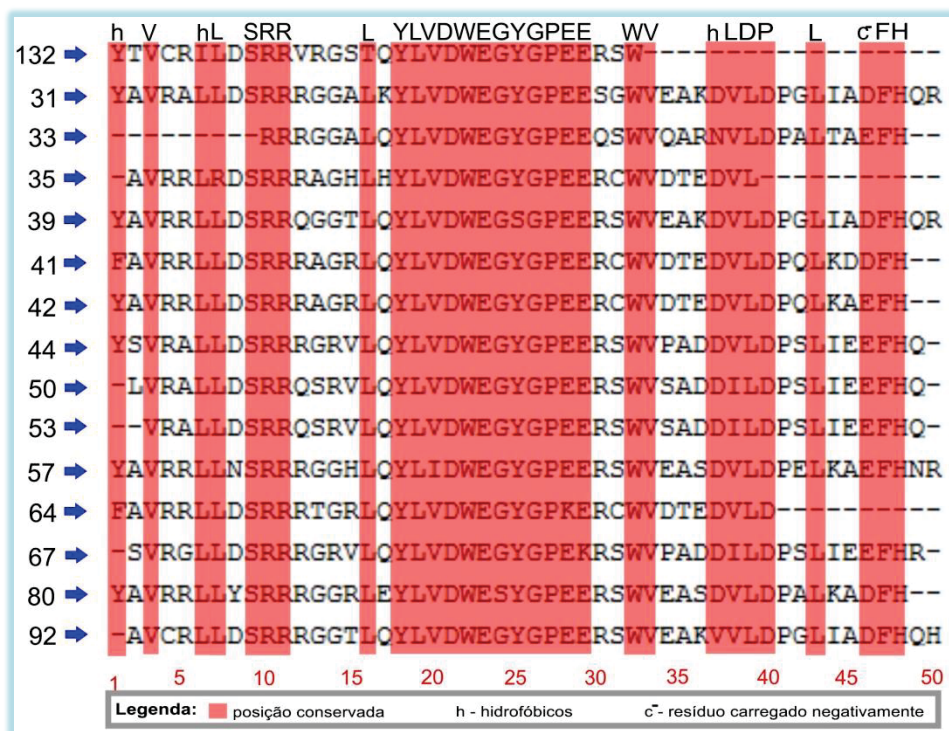


**Supplementary Fig. 7** MUSCLE alignment of Integrase sequences showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of *Apareiodon* sp.

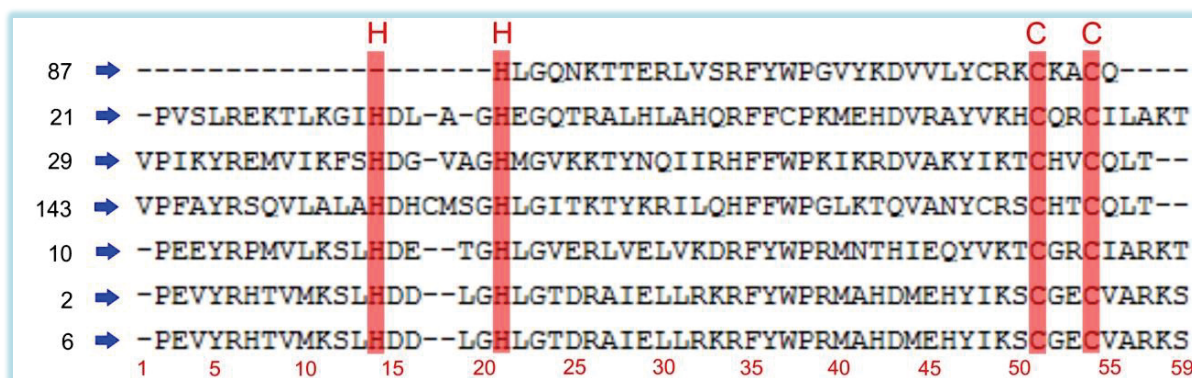




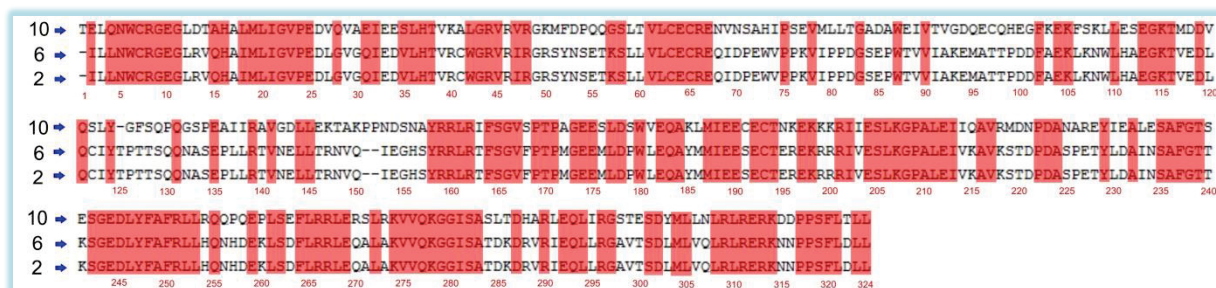
**Supplementary Fig. 8** Alignment MUSCLE of the intact sequences of the Chromo domain evidencing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of *Apareiodon* sp.



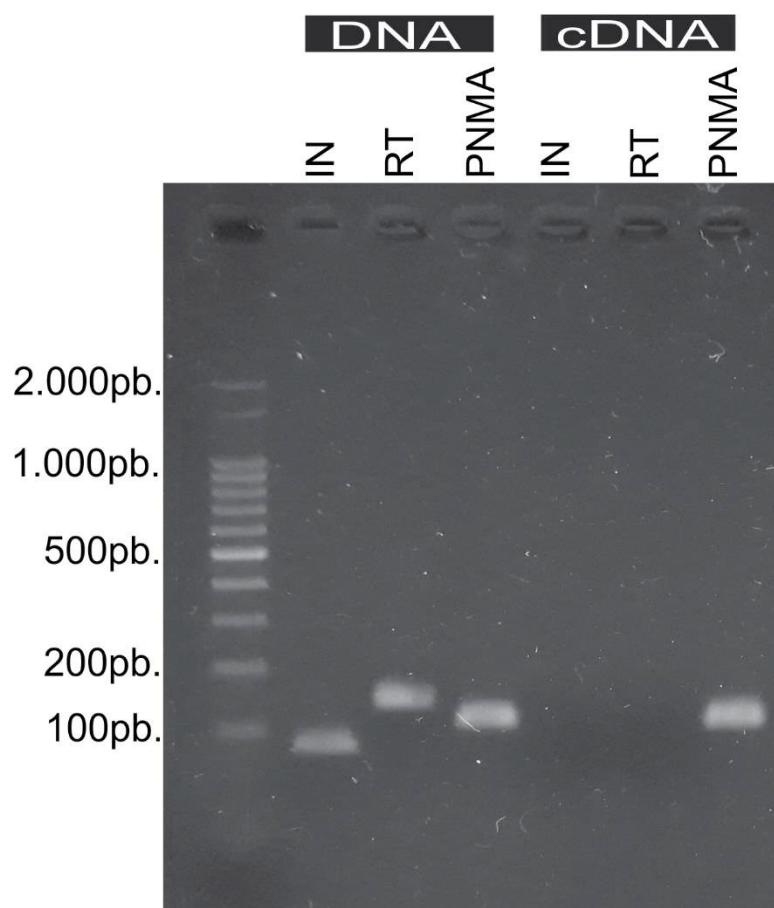
**Supplementary Fig. 9** Alignment MUSCLE of the intact Zf - HHCC sequences showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of *Apareiodon* sp.



**Supplementary Fig. 10** Alignment MUSCLE of the PNMA sequences showing the conserved positions (red marking). The numbers at the bottom of the image refer to the position of the residue in the alignment and the numbers in the left column refer to the sequence identification of *Apareiodon* sp.



**Supplementary Fig. 11** Agarosis electrophoresis gel 3% showing IN, RT and PNMA amplification of *Apareiodon* sp. using genomic DNA (left) and, cDNA of testis (right).



## 5.2 CAPÍTULO 2

### ELEMENTO MÓVEL EN/SPM SENESCENTE EM UM GENOMA DE PEIXE NEOTROPICAL: GÊNESE DE REPETIÇÕES IN TANDEM E MECANISMOS MOLECULARES NA DIFERENCIAÇÃO DOS CROMOSSOMOS SEXUAIS ZW

#### RESUMO

A anotação da fração repetitiva do genoma de *Apareiodon* sp. (sistema de cromossomos sexuais ZZ/ZW) foi realizada recentemente. O estudo revelou uma fase de invasão dos transposons de DNA Tc1-Mariner, En/Spm e Helitron, além de microssatélites na região específica do cromossomo W. Contudo, pouco se sabe sobre os mecanismos moleculares e genômicos envolvidos com a presença dos elementos transponíveis (TEs) em peixes Neotropicais. Este estudo apresentou uma análise genômica e de caracterização molecular detalhada do elemento En/Spm em *Apareiodon* sp., englobando a sua localização cromossômica em espécies de Parodontidae. Nossos dados corroboram a proposição da participação do TE En/Spm na origem do cromossomo sexual heteromórfico W neste grupo. As cópias do En/Spm de *Apareiodon* sp. são degeneradas e sem possibilidade de movimentação como elemento transponível no genoma. No entanto, as análises de predição de domínios proteicos e a verificação de uma abundância de cópias de repetições in tandem em contigs contendo En/Spm sugerem outros mecanismos moleculares derivados deste TE no genoma. Este estudo propõe a gênese dos microssatélites (TA)<sub>n</sub> e (GA)<sub>n</sub> internos as sequências En/Spm com provável função na heterocromatinização do cromossomo W. Análises genômicas dos clusters contendo En/Spm e os sinais de localização cromossômica sugerem também um processo de formação de DNA satélite a partir deste elemento transponível. Os sinais compartilhados da localização *in situ* das sondas En/Spm, (TA)<sub>15</sub> e (GA)<sub>15</sub> em espécies com proto cromossomos sexuais, com cromossomos ZW e, com cromossomos sexuais ZW<sub>1</sub>W<sub>2</sub> ajudam a explicar a origem destes sistemas de cromossomos sexuais no grupo. Todos os dados obtidos demonstram a senescência do En/Spm no genoma e seu envolvimento em novos mecanismos moleculares para a composição da cromatina, especialmente àqueles relacionados aos cromossomos sexuais heteromórficos.

**Palavras-chave:** co-opção molecular, DNAs satélites, elementos transponíveis, genômica, heterocromatina, Parodontidae.

## 1 INTRODUÇÃO

A obtenção de dados genômicos utilizando a tecnologia de sequenciamento em larga escala e análises em bioinformática têm aumentado substancialmente entre os Teleostei (Aparicio et al. 2002; Valente et al. 2014; Coan & Martins 2018). Entre os peixes Neotropicais estes dados são escassos e impossibilitam a comparação mais aprofundada a nível genômico para a análise da extensa diversidade existente no grupo. Dados genômicos foram obtidos para uma espécie de *Apareiodon* (Characiformes: Parodontidae) detentora de cromossomos sexuais heteromórficos Z e W (Schemberger et al. submetido para publicação). Nesta espécie, 36% do genoma é composto de DNAs repetitivos, sendo os DNA transposons e microssatélites os mais abundantes (Schemberger et al. submetido para publicação).

Os peixes Parodontidae, agrupados nos gêneros *Parodon*, *Saccodon* e *Apareiodon*, apresentam espécies caracterizadas citogeneticamente em uma série para a diversificação dos cromossomos sexuais: sem cromossomos sexuais heteromórficos, com proto-cromossomos sexuais e, com sistemas de cromossomos sexuais heteromórficos ZZ/ZW ou ZZ/ZW<sub>1</sub>W<sub>2</sub> (Schemberger et al. 2011). Neste grupo, os DNAs repetitivos promoveram eventos de remodelagem cromossômica/genômica (Vicente et al. 2003; Bellafronte et al. 2011; Schemberger et al. 2011, 2014, 2016; Ziemniczak et al. 2014; Traldi et al. 2016; Nascimento et al. 2018). No genoma de *Apareiodon* sp., um total de 56 superfamílias de elementos transponíveis (TEs) foram identificadas, sendo os DNA transposons Tc1-Mariner, En/Spm e Helitron, além de expansões microssatélites, os responsáveis pela composição da região heteromórfica do cromossomo W (Schemberger et al. submetido para publicação).

TEs são capazes de se mover e replicar contribuindo assim para o tamanho, estrutura e diversidade do genoma hospedeiro (Kidwell 2002) e são responsáveis por uma proporção significativa da variação cariotípica entre espécies (Eichler & Sankoff 2003; Longo et al. 2009; Farré et al. 2011; Glugoski et al. 2018). Kidwell & Lisch (2001) descreveram que o "ciclo de vida" de um TE no genoma hospedeiro consiste em três fases principais: (i) uma fase de invasão curta na qual ocorre amplificação do número de cópias acompanhada por mutação que torna alguns elementos inativos, (ii) um estágio de maturidade no qual a amplificação e a perda de cópias estão mais ou menos em equilíbrio e, (iii) a senescência, a qual se estende por milhões de anos, onde os elementos autônomos são perdidos, não ocorre amplificação e, as



sequências não autônomas deterioram para permanecer neutras ou divergir em função nos genomas. Em relação as funções divergentes nos genomas, Miller et al. (1997) propuseram a “domesticação molecular”, onde cópias deterioradas de TEs podem adquirir funções benéficas ao genoma (para uma revisão ver, Feschotte 2008).

Em perspectivas moleculares, os sítios de ocorrência de TEs podem se envolver em uma série de mecanismos que podem afetar ou influenciar partes do genoma (Kidwell 2002). A recombinação ectotópica entre cópias de um mesmo TE, localizados em regiões genômicas não homólogas, pode levar a rearranjos cromossômicos (Langley et al. 1988). Em outra via, cópias deterioradas de TEs podem dar origem a DNAs satélites para compor a heterocromatina (Dorer & Henikoff 1994; Fanti et al. 1998; Kapitonov & Jurka 2000). Além destes, os TEs podem estar envolvidos em mecanismos de transdução, duplicações gênicas, dispersão de expansões microssatélites (Ramsay 1999), manutenção dos telômeros de *Drosophila* (Okazaki et al. 1995), entre outros (para uma revisão ver Kidwell 2002).

Os TEs foram classificados por Wicker et al. (2007) em Classe I ou retrotransposons, quando utilizam um intermediário de RNA para sua transposição e; os de Classe II ou transposons DNA, que adotaram o mecanismo de transposição via intermediário de DNA, com duas subclasses diferenciadas pelo número de cadeias de DNA que são cortadas durante a transposição. Entre os transposons de DNA, a maioria dos representantes é de subclasse I, e são caracterizados por apresentar duplicação de sítio alvo (TSD), repetições terminais invertidas (TIRs) e uma região codificadora para uma transposase específica (DDD ou DDE) necessária para sua transposição (Wicker et al. 2007). Com base na diversidade estrutural das transposases, os transposons de DNA são agrupados em várias superfamílias, entre elas Tc1-Mariner, Merlin, Piggy-Bac, Hat, Harbinger, P, Transib, Mutator e CACTA (Wicker et al. 2007; Kapitonov & Jurka 2008).

O Enhancer/Supressor mutador (En/Spm) é o elemento pertencente à superfamília CACTA mais estudado (Altinkut et al. 2006; Gbadegesin & Beeching 2010; Kolade et al. 2015; Nouroz et al. 2017). Foram descobertos como elemento autônomo Enhancer (En) (Peterson 1953) e, por McClintock (1954) como Supressor-mutador (Spm) (Pereira & Saedler 1989). O En/Spm está organizado em dois domínios proteicos: (i) *TnpD* que codifica para TNPD ou transposase, responsável pela excisão e integração do TE no genoma e, (ii) *TnpA* codificando para a proteína TNPA, que está envolvida principalmente em processos de dimerização e ligação ao DNA

(Trentmann et al. 1993; Gierl 1996; Gbadegesin & Beeching 2010). Flanqueando a região de codificação, o En/Spm apresenta a região sub-terminal, que abriga os motivos de ligação para TNPA e, nas suas extremidades, as repetições invertidas terminais (TIRs) de 10-28 pb, essenciais para o reconhecimento do local excisão das fitas pela transposase (Wicker et al. 2007; Gbadegesin & Beeching 2010).

A estrutura e função da En/Spm é bem conhecida em plantas (Gbadegesin & Beeching 2010). Neste grupo, estes DNAs transponíveis foram descritos para apresentar mecanismos epigenéticos de metilação envolvidos em silenciamento gênico (Cuypers et al. 1988; Gbadegesin & Beeching 2010). Em *Arabidopsis thaliana*, famílias En/Spm-like originaram o DNA satélite paracentromérico ENSAT1 (Kapitonov & Jurka 1999). No entanto, a estrutura e possíveis mecanismos moleculares de atuação ou influência do elemento En/Spm são precariamente compreendidos em peixes. Steinemann & Steinemann (1998) apresentaram evidências consistentes que a primeira etapa na diferenciação de um cromossomo sexual é guiada pelo acúmulo de TEs e, que este enriquecimento tornaria a estrutura da cromatina de eucromática para heterocromática nestas regiões. Neste contexto e com o conhecimento da localização cromossômica de elementos En/Spm no cromossomo W de *Apareiodon* sp. (Schemberger et al. submetido para publicação) este estudo teve por objetivo a caracterização molecular das cópias En/Spm para avaliar mecanismos moleculares envolvidos com a diversificação genômica deste TE em *Apareiodon* sp. e, na organização cromossômica em diferentes espécies de Parodontidae.

## 2 MATERIAL E MÉTODOS

### 2.1 SELEÇÃO E CARACTERIZAÇÃO DE SEQUÊNCIAS EN/SPM

A fração repetitiva do genoma de *Apareiodon* sp. foi anotada utilizando o software RepeatMasker (Smit et al. 2013-2015) e genoma de referência *Danio rerio* por Schemberger et al. (submetido para publicação). A partir desta biblioteca, as sequências identificadas como CMC-En/Spm foram selecionadas e os contigs com similaridade ao elemento En/Spm foram recuperados do genoma de *Apareiodon* sp. (disponível em <http://sacibase.ibb.unesp.br/jbrowse/JBrowse-1.12.1/index.html?data=data-apareiodon-sp>).

Um script linguagem php foi desenvolvido para a seleção das sequências En/Spm nos contigs de *Apareiodon* sp. As sequências En/Spm obtidas foram submetidas ao software online CENSOR para validação de similaridade (Kohany et al. 2006). Para a identificação das TIRs, e regiões subterminais as sequências En/Spm foram submetidos à análise no software DetectIR (Ye et al. 2014) e também à busca manual pelas sequências 5' terminal CACTGTAAAA e CACAGCAAAA conforme encontrado por Kapitonov & Jurka (2008) em *Danio rerio*.

Posteriormente, as sequências En/Spm foram submetidas a identificação das fases abertas de leitura (ORFs) e a tradução em sequências de aminoácidos com auxílio do banco de dados online FishTEDB (Shao et al. 2018). As sequências similares ao En/Spm foram submetidas ao Blastx e Blastp do NCBI e ao matching Pfam families (El-Gebali et al. 2019) para a identificação dos domínios proteicos.

## 2.2 CONSTRUÇÃO DO REPEAT LANDSCAPE DE EN/SPM

As sequências do elemento En/Spm selecionadas foram executadas no RepeatMasker com os parâmetros “slow (-s)”, “align (-a)”, e “library (-lib)”. Os arquivos resultantes foram submetidos aos scripts Perl auxiliares do Repeat Masker: “createrepeatlandscape.pl” e “calcdivergencefromalign.pl” para cálculo dos valores de divergência de Kimura e construção do Repeat Landscape para o elemento En/Spm.

## 2.3 CONTAGEM DE SEQUÊNCIAS MICROSSATÉLITES NAS SEQUÊNCIAS EN/SPM

As sequências de microsatélites dinucleotídicos (TA)<sub>n</sub> e (GA)<sub>n</sub> foram mensuradas a partir de 5 repetições *in tandem* até o número máximo de repetições destes microsatélites presente nas sequências En/Spm. Para este levantamento foi desenvolvido um script de linguagem em php para leitura de sequências e contabilização da frequência dos microsatélites.

## 2.4 AVALIAÇÃO DE REPETIÇÕES IN TANDEM EM CONTIGS CONTENDO SEQUÊNCIAS EN/SPM

O número de sequências En/Spm em cada contig foi mensurado analisando os resultados do RepeatMasker. Os dados foram filtrados e selecionados os contigs com mais de 10 sequências En/Spm. Posteriormente, para avaliar a presença das repetições estes contigs foram submetidos ao Dotplot do software Geneious 4.8.5 (Drummond et al. 2009) com mecanismo de comparação baseado na ferramenta Dottup. Para a representação gráfica adotamos o sistema clássico de cores, com o azul representado as sequências curtas (< 100 pb) e vermelho para fragmentos maiores que 100 pb.

## 2.5 OBTENÇÃO DE SONDAS PARA A LOCALIZAÇÃO IN SITU

O DNA genômico foi obtido a partir do fígado dos exemplares de *Apareiodon* sp. utilizando brometo de cetil trimetilamônio (CTAB) de acordo com o protocolo adaptado de Doyle e Doyle (1990). Contigs do genoma de *Apareiodon* sp. contendo as sequências En/Spm foram alinhados no software MUSCLE (Edgar 2004) para a detecção de regiões de degeneração mínima. Estas regiões foram utilizadas para o desenho de primers para um segmento En/Spm (livre das expansões microssatélites TA e CAA) utilizando o software Primer3plus (Untergasser & Nijveen 2007; Rozen & Skaletsky 2000). Os melhores primers obtidos foram: Enspm\_fw 5' - gcaggctttcacgaatct - 3' e, Enspm\_rv 5' - catgaggggctcctgttct - 3', com tamanho de fragmento esperado ~510 pb.

A reação de amplificação do En/Spm por PCR utilizou 40 ng de DNA genômico, 0,2 µM de cada primer, 0,4 mM de desoxinucleotídeos trifosfatos (dNTPs), 1× Taq tampão de reação (200 mM Tris, pH 8.4, 500 mM KCl), 2 mM MgCl<sub>2</sub> e 1 U de Taq DNA polimerase (Invitrogen). Os parâmetros da amplificação foram: desnaturação inicial 5 min a 94 °C, 35 ciclos de três passos (1 min a 94 °C, 40 segundos a 56 °C e 45 segundos a 72 °C) e extensão final de 7 min a 72 °C. Os produtos de PCR resultantes foram purificados utilizando o kit PCR DNA and Gel Band Purification (GE Healthcare), submetidos ao sequenciamento automático de DNA utilizando sequenciador de DNA ABI-PRISM Genetic Analyzer (Applied Biosystems) e confirmação de identidade do segmento obtido no software online CENSOR. Confirmada a identidade, a sequência resultante foi marcada com digoxigenina-11-dUTP utilizando o Kit Dig-Nick Translation Mix (Roche Applied Science) para geração de sonda En/Spm e posterior uso em procedimento de localização *in situ*.



Em paralelo, as expansões microssatélites (TA)<sub>n</sub> e (GA)<sub>n</sub> detectadas abundantes nas sequências En/Spm foram marcadas com biotina na extremidade 5' durante sua síntese (Sigma-Aldrich, St. Louis, Missouri, EUA) e também foram utilizadas como sondas na localização *in situ*.

## 2.6 PREPARAÇÕES CROMOSSÔMICAS E LOCALIZAÇÃO IN SITU

Preparações cromossômicas foram obtidas de cinco espécies pertencentes à Parodontidae: *Apareiodon piracicabae* (sem cromossomos sexuais heteromórficos), *Parodon nasus* (proto-cromossomos sexuais) *Apareiodon* sp. (ZZ/ZW), *Parodon hilarii* (ZZ/ZW) e *Apareiodon affinis* (ZZ/ZW<sub>1</sub>W<sub>2</sub>) (Tabela Suplementar 1). A obtenção dos cromossomos mitóticos seguiu o método descrito em Bertollo et al. (2015). Os procedimentos deste estudo estão de acordo com o Comitê de Ética do Uso de Animais da Universidade Estadual de Ponta Grossa (Protocolo: 11/2017).

As sondas En/Spm, (TA)<sub>15</sub> e (GA)<sub>15</sub> foram utilizadas na hibridação *in situ* fluorescente (FISH) de acordo com o descrito em Pinkel et al. (1986). A mistura de hibridação foi sintetizada com as seguintes condições: 2.5 ng/μL de sonda, 50% formamida, 2x SSC, e 10% sulfato dextrano. A hibridação ocorreu por 18 h à 42 °C. Banhos de pós-hibridização foram realizados nas condições: 50% de formamida/2xSSC, 42 °C por 20 min, 0.1x SSC, 60 °C por 15 min, e 4x SSC 0.05% Tween em temperatura ambiente por 10 min. Para a detecção de fluorescência foram utilizados 4 μg/mL de Streptavidin Alexa Fluor 488 (Molecular Probes, Carlsbad, CA, USA) e 10 μg/mL de anti-digoxigenin rhodamine (Roche Applied Science), os quais foram incubados a 37 °C por 1 hora no tampão de detecção [4x SSS, 0.2% (v/v) Tween 20, 5% (w/v) de leite em pó desnatado]. Os cromossomos foram contracolorados com 0,2 μg/mL de 4', 6-diamidino-2-fenilindole (DAPI) em meio de montagem Vectashield (Vector, Burlingame, CA) e analisados em microscópio de campo claro e epifluorescência Zeiss Axio Lab 1 acoplado ao software Zen Lite em Câmera Zeiss AxioCam ICM1, resolução de 1.4 megapixels (Carl Zeiss®).

As metáfases foram organizadas em cariótipos usando o software Adobe Photoshop, versão CS6 e classificados em metacêntricos (m), submetacêntricos (sm), subtelocêntricos (st) e acrocêntricos (a) conforme proposto por Levan et al. (1964).

### 3 RESULTADOS

#### 3.1 IDENTIFICAÇÃO, SELEÇÃO E CARACTERIZAÇÃO DO ELEMENTO EN/SPM

Os dados obtidos a partir da análise no RepeatMasker identificaram um total de 49.560 sequências no genoma macho e, 53.284 no genoma fêmea, todas com similaridade mínima de 65% à superfamília CMC-En/Spm de *Danio rerio*. Estes valores representam 1,44% e 1,56% da fração repetitiva dos genomas macho e fêmea, respectivamente. Após a filtragem obtivemos 18.276 sequências com similaridade apenas ao elemento En/Spm no genoma macho, com tamanho variando de 10 a 3.868 pb e, 19.985 sequências no genoma fêmea, com fragmentos de 10 a 3.179 pb.

As duas estratégias utilizadas para a detecção das TIRs evidenciaram a degeneração desta região nas sequências En/Spm em *Apareiodon* sp., não sendo possível detectar TIRs perfeitas. O software DetectIR identificou TIRs com no máximo 4pb e não apresentaram similaridade desta região entre as sequências, sugerindo não ser uma região TIR e sim uma região palindrômica ao acaso (dados não mostrados). Na busca manual destas regiões utilizando sequências consensos das TIRs de *Danio rerio* detectamos a região da TIR conservada em apenas uma das extremidades das sequências, sendo a outra extremidade altamente degenerada (Fig. Suplementar 1). Nesta análise detectamos o consenso 5'-CACAGCAAAA-3' da TIR em 13 sequências En/Spm do genoma macho e, 22 no genoma fêmea (Fig. Suplementar 1a, b). O consenso 5'-CACTGTAAAA-3' da TIR foi encontrado em 107 sequências En/Spm no genoma macho e 109 sequências no genoma fêmea (Fig. Suplementar 1c e d). Assim, das 38.261 sequências analisadas apenas 0,0056% apresentam a TIR conservada em uma das extremidades. Não foi possível detectar similaridade para as regiões subterminais das sequências En/Spm analisadas.

Para a região codificadora não foram detectadas ORFs para TNPA. Para TNPD foram detectadas cinco sequências com o domínio da transposase, porém todos se mostraram truncados, sendo duas no genoma macho e três no genoma fêmea. Nesta análise, a região interna similar ao elemento En/Spm apresentou 46 sequências que tiveram a predição de 11 tipos de domínios proteicos íntegros, não característicos do elemento En/Spm, sendo eles: Endomucin, GNT-I, HSP90, Keratin

B2, Laminin II, LbR like superfamília, Myristoyl-CoA, Pentapeptide, S-AdoMet Synt\_C, TIP49 P-loop e Zinc finger-C<sub>2</sub>H<sub>2</sub> (Tabela 1).

### 3.2 ANÁLISE GRÁFICA DA INVASÃO DO GENOMA PELO TRANSPOSON EN/SPM

A representação gráfica *landscape* indicou apenas uma onda de invasão do elemento En/Spm durante a evolução do genoma de *Apareiodon* sp. Os gráficos demonstram que a maioria das sequências En/Spm apresentam um alto nível de substituição nucleotídica nos genomas macho e fêmea (Fig. 1). No genoma macho foi detectado que o maior número das cópias En/Spm apresentam 14 a 28% de substituição nucleotídica (Fig. 1a). No genoma fêmea, o maior número de cópias apresenta 16 a 32% de degeneração mutacional (Fig. 1b). O número de cópias En/Spm com níveis de substituição nucleotídica próximo a 0 foi extremamente baixo, tanto na análise do genoma fêmea, quanto na análise do genoma macho (Fig. 1).

### 3.3 IDENTIFICAÇÃO DE SSR (GA/CT)<sub>n</sub> E (TA/AT)<sub>n</sub> NAS SEQUÊNCIAS EN/SPM DE *Apareiodon* sp.

A investigação da presença e número de cópias dos microssatélites internos nas sequências En/Spm de *Apareiodon* sp. resultou em 1.948 *clusters* GA/CT no genoma macho e 4.401 *clusters* no genoma fêmea, os quais variaram em tamanho de cinco para 44 repetições *in tandem* (Tabela 2). Para o microssatélite TA/AT foram encontrados 6.251 *clusters* no genoma macho e 6.561 *clusters* no genoma fêmea, com variação de tamanho de cinco a 26 repetições *in tandem* (Tabela 2). A expansão microssatélite GA/CT é 128% mais frequente no genoma fêmea quando comparado ao genoma macho. Já o microssatélite TA/AT é 5% mais frequente no genoma fêmea quando comparado ao genoma macho.

### 3.4 PRESENÇA DE REPETIÇÕES IN TANDEM EM CONTIGS CONTENDO SEQUÊNCIAS EN/SPM

Na avaliação individual de cada contig foi possível detectar que 3.251 contigs no genoma macho e 3.328 contigs no genoma fêmea abrigavam mais de uma sequência En/Spm. Após a filtragem por número de repetição En/Spm maior que 10

foram obtidos 36 contigs no genoma macho apresentando 10 a 29 repetições En/Spm e, 25 contigs no genoma fêmea contendo 10 a 22 repetições En/Spm (Tabela Suplementar 2 a,b, respectivamente). Estes 61 contigs foram utilizados como *input* para a representação gráfica de DotPlot. Duas sequências no genoma macho (contigs 1553664 e 1610658) e três no genoma fêmea (contigs 1031971, 235765 e 3250488) apresentaram uma densidade expressiva de fragmentos curtos e longos de repetições *in tandem* (Fig. Suplementar 2).

### 3.5 AMPLIFICAÇÃO, SEQUENCIAMENTO E ANÁLISE DE SIMILARIDADE DA SEQUÊNCIA EN/SPM

O produto da amplificação por PCR resultou em um segmento de 511 pb. A análise deste segmento após o sequenciamento automático apresentou similaridade de 69,28% com o elemento En/Spm de *Danio rerio* (número de acesso BX005001).

### 3.6 HIBRIDIZAÇÃO IN SITU FLUORESCENTE

A aplicação da FISH com a sonda de En/Spm revelou a presença deste elemento disperso ao longo dos autossomos, com acúmulos nas regiões centroméricas e terminais dos braços q de alguns cromossomos, e nos cromossomos sexuais de *Apareiodon* sp., *Parodon hilarii* e *Apareiodon affinis* (Fig. 2). Em *A. piracicabae* (sem heteromorfismo de cromossomos sexuais) sítios En/Spm foram verificados na região proximal do par cromossômico 1, além de sinais dispersos ao longo de vários cromossomos (Fig. 2a). Em *P. nasus* foram visualizados pequenos acúmulos da sequência En/Spm nas regiões terminais e subterminais de alguns cromossomos e, na região proximal dos pares cromossômicos 5, 13 (protocromossomo sexual) e 14 (Fig. 2b).

Em *Apareiodon* sp. a sequência En/Spm foi localizada abundante na região q terminal do par 11, na região proximal do par 6 e cromossomo Z, na região intersticial do cromossomo W, além de outros sinais dispersos em alguns cromossomos (Fig. 2c). Em *P. hilarii* marcações En/Spm foram localizadas em regiões terminais e proximais de inúmeros autossomos, na região proximal do cromossomo Z e na região intersticial do braço q do cromossomo W (Fig. 2d). Na espécie *A. affinis* a sequência En/Spm foi localizada em sinais dispersos principalmente nos autossomos 4, 8, 9, 16,

17 e 19, no cromossomo Z (par 1) e nos cromossomos sexuais  $W_1$   $W_2$ , sendo em  $W_1$  localizado na região q subterminal e em  $W_2$  com sinais dispersos ao longo de todo o cromossomo (Fig. 2e).

A localização *in situ* dos microssatélites (GA)<sub>15</sub> e (TA)<sub>15</sub> evidenciaram marcações nas regiões terminais dos braços cromossômicos nas cinco espécies analisadas (Fig. 3). Além destes, *A. piracicabae* apresentou sítios (TA)<sub>n</sub> adicionais nas regiões intersticiais colocalizados com DNA ribossômico 45S (Fig. 3a). Para as espécies com cromossomos sexuais, (GA)<sub>n</sub> está acumulado no braço q dos cromossomos W de *Apareiodon* sp. e *P. hilari* (Fig. 3c, d). Já em *A. affinis* é possível visualizar um maior acúmulo no braço p do cromossomo Z, um pequeno acúmulo na região p terminal do cromossomo Z, além de acúmulos terminais nos demais cromossomos (Fig. 3e). O microssatélite (TA)<sub>n</sub>, por sua vez, apresentou um acúmulo na região proximal dos cromossomos ZW, além das marcas terminais em todos os cromossomos em *Apareiodon* sp. (Fig. 3c). Em *P. hilarii* o microssatélite (TA)<sub>n</sub> é visualizado em toda a extensão do braço q do cromossomo W, além de sítios terminais em vários autossomos (Fig. 3d). Em *A. affinis* o (TA)<sub>n</sub> foi detectado em acúmulos no braço p e na região terminal do braço p do cromossomo Z, além de sítios terminais e todos os cromossomos (Fig. 3e).

#### 4 DISCUSSÃO

Elementos transponíveis compõem uma parte significativa dos principais genomas eucarióticos e desempenham papéis fundamentais na regulação gênica e evolução dos genomas (Slotkin 2018). Nos últimos anos diversos scripts e ferramentas computacionais foram desenvolvidos para a identificação de TEs, mas ainda é um desafio identificá-los e caracterizá-los devido às suas diversas modificações estruturais ou degenerações moleculares (Gao et al. 2014; Nouroz et al. 2015; Schemberger et al. 2016). Este estudo é pioneiro na caracterização do elemento de classe II En/Spm em peixes neotropicais e abre perspectivas para investigações deste TE na diversificação genômica do grupo. Os dados deste estudo mostraram uma deterioração mutacional do elemento En/Spm levando a modificações moleculares e envolvimento em outros mecanismos genéticos e/ou classes de cromatina no genoma de *Apareiodon* sp.

#### 4.1 DETERIORAÇÃO MOLECULAR DO ELEMENTO EN/SPM NO GENOMA DE *Apareiodon* sp.

Os transposons de DNA são os principais componentes da fração repetitiva do genoma de peixes teleósteos, representando 72,7% em *Danio rerio*, 19,4% em *Oryzias latipes* (Gao et al. 2016), 38,7% em *Astatotilapia latifasciata* (Coan & Martins 2018) e 37,9% em *Apareiodon* sp. (Schemberger et al. submetido para publicação). Devido a sua dinâmica de transposição é sugerido que os elementos de classe II estejam envolvidos em diversos mecanismos genéticos ainda muito pouco compreendido em peixes (Gao et al. 2016). Os dados deste estudo demonstraram que elemento En/Spm detém maior número de cópias no genoma fêmea, está relacionado a diferenciação do cromossomo sexual W e, apresenta alto nível de degeneração molecular de sua identidade no genoma.

A análise das sequências nucleotídicas após filtragem para apenas elementos de similaridade ao En/Spm demonstraram sequências com comprimento máximo de ~3.8kb em *Apareiodon* sp., inferior ao ~8.2 kb do En/Spm de plantas (Pereira et al., 1986). Uma vez que elementos En/Spm são constituídos de TIRs em ambas extremidades, regiões sub-terminais flanqueando as TIRs e ORFs para as sequências TnpD e TnpA (Trentmann et al. 1993; Gierl 1996; Gbadegesin & Beeching 2010) estes segmentos foram prospectados nas sequências En/Spm dos genomas macho e fêmea de *Apareiodon*. As TIRs do TE En/Spm apresentam tamanho variável entre as espécies (Trentmann et al. 1993; Kapitonov & Jurka 2008). Em nosso estudo buscamos as TIRs 5' e 3' com o software DetectIR e com base naquelas encontradas no elemento En/Spm do genoma de referência *Danio rerio* (Kapitonov & Jurka 2008). A análise demonstrou a ausência de TIRs conservadas em ambas as extremidades das sequências. Quando presentes, as TIRs foram encontradas conservadas em apenas uma das extremidades. Na maioria das cópias En/Spm de *Apareiodon* sp. as TIRs apresentaram alta taxa de substituição nucleotídica, além de inserções e deleções (*Indels*). As TIRs são regiões para o reconhecimento pela enzima transposase no corte das fitas de DNA e excisão do elemento (Capy et al. 1998; Wicker et al. 2007). A ausência destes segmentos indica inatividade de En/Spm no genoma de *Apareiodon*.

Além das TIRs foram realizadas buscas às regiões subterminais do elemento En/Spm no genoma de *Apareiodon* sp. e as análises não retornaram segmentos

similares. As regiões subterminais são descritas a atuar no controle da transcrição deste elemento e exclusões ou mutações nesta região diminuem ou inibem os eventos de transposição (para revisão ver Gierl et al. 1985; Schwarz-Sommer et al. 1985; Tacke et al. 1986). A ausência desta região é mais um indicativo da ausência de mobilização do En/Spm no genoma de *Apareiodon*.

Corroboram com a proposição de inatividade do elemento En/Spm a ausência de domínios proteicos íntegros para TNPA (domínio de ligação às TIRs) e TNPD (transposase) no genoma de *Apareiodon* sp. Análise de similaridade destas regiões também demonstraram alta taxa de polimorfismos (SNP) e *Indels*. Estes dados também são evidenciados pela análise gráfica de *landscape* do elemento En/Spm no genoma de *Apareiodon* sp., que demonstrou uma única onda de invasão antiga do En/Spm na linhagem de *Apareiodon* sp. e uma alta taxa de degeneração da maioria de suas cópias no genoma. Estes dados estão de acordo com o proposto por Schemberger et al. (submetido para publicação), os quais verificaram duas ondas principais de invasão genômica de elementos transponíveis em *Apareiodon* sp., estando o elemento En/Spm na onda de invasão antiga e, segundo estes autores, que deu origem a diferenciação do cromossomo W.

Na análise *landscape* é possível verificar também que as cópias En/Spm estão com degeneração maior no genoma de fêmea. Esta degeneração maior pode ser consequência da presença deste TE durante a etapa de diferenciação da região repetitiva do cromossomo W. Os mecanismos atuantes na diferenciação das sequências presentes na região repetitiva dos cromossomos sexuais são precariamente compreendidos em peixes. No geral, são conhecidos os mecanismos propostos em *Drosophila*, onde após a invasão de TEs no cromossomo sexual heteromórfico, estes degeneram e tendem a sofrer heterocromatização (Steinemann & Steinemann 1998). Os TEs apresentam um ‘ciclo de vida’ no genoma (Kidwell & Lisch 2001). Nossos dados demonstram que o elemento En/Spm está na fase de senescência do seu ‘ciclo de vida’, onde são observadas a perda dos elementos autônomos, incapacidade de transposição, degeneração das sequências, sua permanência como fósseis neste genoma ou, sua possibilidade de atuar em outros mecanismos na cromatina ou, para a co-opção molecular de novas funções no genoma (Kidwell & Lisch 2001; Kidwell 2002; Feschotte 2008).



## 4.2 CO-OPÇÃO MOLECULAR DO ELEMENTO EN/SPM

O conceito de domesticação, co-opção ou exaptação molecular foi utilizado para descrever o processo onde uma sequência TE pode fornecer matéria bruta de DNA para geração de novas funções e trazer benefícios para o genoma (Miller et al. 1997). Entre as novas funções que podem ser exercidas pelos TEs co-optados estão a geração de novos sítios regulatórios da expressão gênica (TFBS – *transcription factor binding sites*), geração de sítios isoladores, formar novos genes, geração de micro RNAs, diferenciar em proteínas de ligação ao DNA (DBPs – *DNA binding proteins*) envolvidas com a regulação gênica e divisão celular, entre outros (Feschotte & Pritham 2007; Feschotte 2008). A busca por domínios proteicos internos às sequências En/Spm no genoma de *Apareiodon* sp., apesar de não retornar domínios TNPA e TNPD característicos deste elemento móvel, detectaram domínios proteicos íntegros para Endomucin, GNT-I, HSP90, Keratin B2, Laminin II, LbR like superfamília, Myristoyl-CoA, Pentapeptide, S-AdoMet Synt\_C, TIP49 P-loop e Zinc finger-C<sub>2</sub>H<sub>2</sub>. Todas estes domínios apresentam funções muito diferenciadas nos genomas e sua diferenciação a partir de cópias En/Spm permanecem para ser investigadas em uma análise mais profunda de genômica comparativa.

## 4.3 GÊNESE DE SEQUÊNCIAS MICROSSATÉLITES INTERNAS AO ELEMENTO EN/SPM

Dados obtidos neste estudo demonstraram uma grande abundância dos microsatélites (GA)<sub>n</sub> e (TA)<sub>n</sub> internos as sequências de similaridade ao En/Spm, com maior frequência no genoma fêmea. Alguns mecanismos moleculares são associados a presença de TEs e microsatélites. Transposons de DNA, a exemplo do Tc1-Mariner foi descrito a apresentar uma maior facilidade de inserção em regiões ricas TA (Plasterk et al. 1999). Em outra via, a dispersão dos microsatélites nos cromossomos é proposta a ocorrer como efeito carona durante a movimentação de elementos transponíveis (Coates et al. 2010). Wilder & Hollocher (2001) descreveram que TEs contendo regiões proto-microsatélites comumente podem se expandir em comprimento e promover a gênese de repetições microsatélites. Adicionalmente, o mecanismo de *DNA slippage* tem sido descrito para gerar a variação do número de repetições em microsatélites pré-existent (Levinson & Gutman 1987).



Os dados deste estudo demonstraram a colocação dos microssatélites (TA)<sub>n</sub> e (GA)<sub>n</sub> e o elemento En/Spm essencialmente nos sítios presentes nos cromossomos sexuais, com os maiores acúmulos nos cromossomos W de *Apareiodon* sp., *P. hilarii* e ZW<sub>1</sub>W<sub>2</sub> de *A. affinis*. Estes dados refutam a dispersão destes dois microssatélites nos cromossomos das espécies de Parodontidae avaliadas como consequência da movimentação do En/Spm. Assim, uma vez que o En/Spm foi proposto a estar na fase de senescência e sem mobilização no genoma de *Apareiodon* sp., suas cópias apresentando proto-microssatélites (TA)<sub>n</sub> e (GA)<sub>n</sub> poderiam explicar a gênese destes microssatélites nos cromossomos sexuais de Parodontidae, conforme mecanismo descrito por Wilder & Hollocher (2001). Ao contrário, a presença do En/Spm não explica a abundância dos microssatélites (TA)<sub>n</sub> e (GA)<sub>n</sub> em regiões terminais dos autossomos das espécies de Parodontidae avaliadas.

A densidade de sítios microssatélites é maior nas extremidades dos braços cromossômicos em inúmeras espécies de peixes (Ziemniczak et al. 2014; Scacchetti et al. 2015; Pucci et al. 2016). A densidade, a localização e/ou a distribuição de microssatélites no cromossomo sexual heteromórfico difere daquela encontrada nos autossomos (Bell & Jurka 1997; Ziemniczak et al. 2014; Scacchetti et al. 2015). Esta correspondência de localização de sequências microssatélites na região repetitiva específica do cromossomo sexual heteromórfico foi sugerida a participar do processo de diferenciação e diminuição da frequência de recombinação entre os cromossomos Z e W ou X e Y (Bell & Jurka 1997; Ziemniczak et al. 2014; Scacchetti et al. 2015). No entanto, evidências mais consistentes do grande acúmulo de sequências microssatélites nos cromossomos sexuais heteromórficos advém de sua possibilidade de formar estruturas tríplex (Frank-Kamenetskii & Mirkin 1995) para formação de corpos heterocromáticos no cromossomo W (Saitoh et al. 1991; Komissarov et al. 2018).

Na hipótese de origem e diversificação dos cromossomos sexuais proposta por Steinemann & Steinemann (1998) os elementos transponíveis tendem a se acumular rapidamente para formar a região heteromórfica do cromossomo Y, ou W e, desta forma, restringir a recombinação com seu homólogo X, ou Z. Em um segundo momento, estas áreas expandidas no cromossomo heteromórfico deveriam modificar a estrutura da cromatina de eucromática à heterocromática. Os dados deste estudo demonstram que o elemento En/Spm acumulou na origem da região heteromórfica do

cromossomo W de *Apareiodon* e, que a gênese dos microssatélites (TA)<sub>n</sub> e (GA)<sub>n</sub> podem estar relacionadas a eventos de heterocromatinização.

#### 4.4 PROPOSTA DE GÊNESE DE DNA SATÉLITE A PARTIR DO TE EN/SPM

Os dados de localização *in situ* da sequência En/Spm na espécie *Apareiodon* sp. mostraram sítios cromossômicos consistentes ao padrão encontrado aos DNAs satélites. Vicari et al. (2006) demonstraram a presença de blocos heterocromáticos para as regiões q terminal do par 11, na região proximal do par 6 e do cromossomo Z e, no braço q do cromossomo W em *Apareiodon* sp. Estes dados em conjunto com as análises deste estudo dos grafos do Repeat Explorer que contém En/Spm e a avaliação do padrão de repetição dos *clusters* En/Spm com a ferramenta DotPlot demonstraram a ocorrência de unidades de tamanhos maiores e menores que 100 pb, tanto no genoma macho, quanto no genoma fêmea, os quais estariam em densidade comum às repetições *in tandem*, conforme proposto em Maizel & Lenk (1981).

Estes contigs podem ser correlacionados aos que compõem os blocos heterocromáticos presentes nos autossomos 6 e 11 e àqueles presentes nos cromossomos ZW de *Apareiodon*. Estes resultados explicam também a maior abundância de cópias En/Spm no genoma fêmea devido a sua expansão na região repetitiva do cromossomo W. Alguns DNAs satélites podem ter origem a partir de TEs ou, apresentar um componente maior relacionado a uma parte do elemento móvel (Evgen'ev et al. 1982; Zelentsova et al. 1986; Heikkinen et al. 1995; Kapitonov et al. 1999). Outros estudos descrevem que o DNA satélite é o principal componente da heterocromatina, a qual também é enriquecida por outros DNAs repetitivos, tais como TEs, minissatélites e microssatélites (Mazzuchelli & Martins 2009; Vicari et al. 2010; Schemberger et al. 2014). Em *Arabidopsis thaliana* o DNA satélite ENSAT1 de localização centromérica foi descrito por apresentar 22 repetições *in tandem* de uma unidade de 1949 pb, das quais 499 pb eram 72% similares ao início de um elemento En/Spm denominado Atenspm2 (Kapitonov et al. 1999). Com os resultados obtidos neste estudo é possível inferir que o acúmulo de sequências En/Spm em blocos heterocromáticos de *Apareiodon* sp. indicam a emergência de um DNA satélite no genoma, embora a unidade de repetição, no estágio atual, está fragmentada em vários segmentos de poucos pares de bases.

#### 4.5 DIFERENCIAÇÃO DOS CROMOSSOMOS SEXUAIS

A proposta de origem dos cromossomos sexuais de Parodontidae descreve uma inversão cromossômica realocando segmentos repetitivos terminais para a região proximal do proto cromossomo sexual e, a partir deste sítio proximal a expansão da região heteromórfica de cromossomo W (Schemberger et al. 2011). Análises genômicas comparativas da fração repetitiva de *Apareiodon* sp. demonstraram que a região heteromórfica do cromossomo W foi originada pela invasão dos elementos Tc1-Mariner, En/Spm, Helitron, além de deter expansões microsatélites (Schemberger et al. submetido para publicação). Neste estudo é demonstrado um pequeno acúmulo proximal de En/Spm no proto sexual (par 13) de *P. nasus* e nos cromossomos ZW de *Apareiodon* sp., *P. hilarii* e, ZW<sub>1</sub>W<sub>2</sub> de *A. affinis*, indicando a presença deste TE na origem dos cromossomos sexuais destas espécies.

A colocação dos sítios En/Spm com os microsatélites (TA)<sub>n</sub> e (GA)<sub>n</sub> nos cromossomos sexuais de *Apareiodon* sp., *P. hilarii* e *A. affinis* também é evidente neste estudo. Em atenção, a comparação da localização dos microsatélites (TA)<sub>n</sub> e (GA)<sub>n</sub> nas espécies *Apareiodon* sp. e *Parodon hilarii* demonstram uma composição e abundância diferenciada no cromossomo W. Enquanto o microsatélite (GA)<sub>n</sub> é mais abundante no braço longo do cromossomo W de *Apareiodon* sp., esta situação se inverte em *P. hilarii*, onde a repetição (TA)<sub>n</sub> ocupa a maior extensão do braço longo do W. Esses resultados reforçam a provável gênese destes microsatélites internos às sequências En/Spm em Parodontidae, conforme proposta de Wilder & Hollocher (2001) e, apontam que estas espécies seguiram caminhos evolutivos diferenciados para o acúmulo destas sequências no cromossomo W.

O sistema de cromossomos sexuais múltiplo ZZ/ZW<sub>1</sub>W<sub>2</sub> de *A. affinis* foi proposto ter origem por fissão cêntrica do cromossomo 1 (Z) seguida de inversões pericêntricas para origem do W<sub>1</sub> e W<sub>2</sub> (Moreira-Filho et al. 1980). Posteriormente, Schemberger et al. (2011) utilizaram pintura cromossômica com sonda W parcial e demonstraram uma possível translocação proto cromossomos sexuais e autossomo para origem do sistema ZZ/ZW<sub>1</sub>W<sub>2</sub>. A origem deste sistema múltiplo em *A. affinis* ainda é incerta. No entanto, a colocação do elemento En/Spm e dos microsatélites (TA)<sub>n</sub> e (GA)<sub>n</sub> no cromossomo Z, W<sub>1</sub> e W<sub>2</sub>, assim como nos cromossomos ZW de *Apareiodon* sp. e *Parodon hilarii* reforçam a proposição de

translocação proto cromossomos sexuais e autossomo para a origem do sistema múltiplo ZZ/ZW<sub>1</sub>W<sub>2</sub> de *Apareiodon affinis*.

## 5 CONCLUSÃO

A avaliação das cópias En/Spm no genoma de *Apareiodon* sp. demonstrou que este TE está em fase de senescência. Os dados demonstraram a gênese de microssatélites internos às cópias En/Spm presentes no cromossomo W e, estes provavelmente estão relacionados a estruturas de DNA triplex para formação da heterocromatina. Análises de contigs com abundância de repetições En/Spm sugerem também a provável formação de DNA satélite para compor blocos heterocromáticos. A localização *in situ* do En/Spm corrobora a proposta da presença deste TE na origem do sistema de cromossomos sexuais no grupo. Nosso conjunto de dados demonstrou a emergência de outros mecanismos genéticos de composição da cromatina que surgiram destes elementos En/Spm fósseis no genoma.

## REFERÊNCIAS

As referências utilizadas neste artigo se encontram ao final da tese.

Tabela 1 Predição de domínios proteicos detectados nas sequências que apresentaram similaridade ao elemento En/Spm dos genomas macho/fêmea de *Apareiodon* sp.

Identificação da sequência	Sexo	Domínio	Descrição
3142478	♀	Endomucin	Endomucin Family
2734187	♀	GNT-I	GNT-I family
1130423	♀	HSP90	Hsp90 protein
2294041	♀	Keratin_B2_2	Keratin, high sulfur B2 protein
2889985	♀	Laminin_II	Laminin Domain II
1642025	♀	LbR like superfmilia	LbR like superfmilia
1712438	♀	LbR like superfmilia	LbR like superfmilia
2370492	♀	LbR like superfmilia	LbR like superfmilia
2466757	♀	LbR like superfmilia	LbR like superfmilia
2553158	♀	LbR like superfmilia	LbR like superfmilia
649802	♀	LbR like superfmilia	LbR like superfmilia
666269	♀	LbR like superfmilia	LbR like superfmilia
233783	♀	NMT	Myristoyl-CoA:protein N-myristoyltransferase, N-
2615852	♀	NMT	Myristoyl-CoA:protein N-myristoyltransferase, N-
327849	♂	Pentapeptide	Pentapeptide repeats (8 copies)
1086812	♂	Pentapeptide	Pentapeptide repeats (8 copies)
1205005	♂	Pentapeptide	Pentapeptide repeats (8 copies)
327849	♂	Pentapeptide	Pentapeptide repeats (8 copies)
1086812	♂	Pentapeptide	Pentapeptide repeats (8 copies)
984853	♂	Pentapeptide	Pentapeptide repeats (8 copies)
1473471	♀	Pentapeptide	Pentapeptide repeats (8 copies)
16693	♀	Pentapeptide	Pentapeptide repeats (8 copies)
2298714	♀	Pentapeptide	Pentapeptide repeats (8 copies)
2618369	♀	Pentapeptide	Pentapeptide repeats (8 copies)
316548	♀	Pentapeptide	Pentapeptide repeats (8 copies)
3265251	♀	Pentapeptide	Pentapeptide repeats (8 copies)
639179	♀	Pentapeptide	Pentapeptide repeats (8 copies)
1181029	♀	Pentapeptide	Pentapeptide repeats (8 copies)
1066749	♀	S-AdoMet_synt_C	S-adenosylmethionine synthetase, C-terminal
3446731	♀	S-AdoMet_synt_C	S-adenosylmethionine synthetase, C-terminal
3414452	♀	TIP49	TIP49 P-loop
658114	♀	TIP49	TIP49 P-loop
257690	♂	zf-C <sub>2</sub> H <sub>2</sub>	Zinc finger-C <sub>2</sub> H <sub>2</sub>
2011642	♂	zf-C <sub>2</sub> H <sub>2</sub>	Zinc finger-C <sub>2</sub> H <sub>2</sub>
214897	♂	zf-C <sub>2</sub> H <sub>2</sub>	Zinc finger-C <sub>2</sub> H <sub>2</sub>
1562487	♂	zf-C <sub>2</sub> H <sub>2</sub>	Zinc finger-C <sub>2</sub> H <sub>2</sub>
1315788	♂	zf-C <sub>2</sub> H <sub>2</sub>	Zinc finger-C <sub>2</sub> H <sub>2</sub>
258753	♂	zf-C <sub>2</sub> H <sub>2</sub>	Zinc finger-C <sub>2</sub> H <sub>2</sub>
1636461	♀	zf-C <sub>2</sub> H <sub>2</sub>	Zinc finger-C <sub>2</sub> H <sub>2</sub>
165588	♀	zf-C <sub>2</sub> H <sub>2</sub>	Zinc finger-C <sub>2</sub> H <sub>2</sub>
2372812	♀	zf-C <sub>2</sub> H <sub>2</sub>	Zinc finger-C <sub>2</sub> H <sub>2</sub>
3569593	♀	zf-C <sub>2</sub> H <sub>2</sub>	Zinc finger-C <sub>2</sub> H <sub>2</sub>
3677149	♀	zf-C <sub>2</sub> H <sub>2</sub>	Zinc finger-C <sub>2</sub> H <sub>2</sub>
821708	♀	zf-C <sub>2</sub> H <sub>2</sub>	Zinc finger-C <sub>2</sub> H <sub>2</sub>
924088	♀	zf-C <sub>2</sub> H <sub>2</sub>	Zinc finger-C <sub>2</sub> H <sub>2</sub>

Tabela 2 Tamanho e número de repetições microssatélites GA/CT e TA/AT detectadas nas sequências En/Spm dos genomas macho e fêmea *Apareiodon* sp.

N. de repetições <i>in tandem</i>	Sequências microssatélites			
	GA/CT		TA/AT	
	♂	♀	♂	♀
5	857	1.737	1913	2137
6	335	922	844	881
7	250	519	652	652
8	118	391	674	645
9	102	208	638	637
10	63	164	632	615
11	72	145	520	596
12	34	70	206	206
13	20	59	32	64
14	23	40	39	35
15	14	32	25	23
16	10	27	23	20
17	11	21	14	21
18	11	10	10	6
19	10	10	5	9
20	7	6	6	2
21	1	7	6	2
22	2	4	4	2
23	2	3	4	2
24	1	1	2	1
25	-	5	2	3
26	-	5	1	2
27	-	2	-	-
28	1	-	-	-
30	-	1	-	-
31	-	2	-	-
32	-	2	-	-
33	-	1	-	-
34	1	1	-	-
35	-	1	-	-
36	-	1	-	-
38	1	1	-	-
39	-	2	-	-
40	1		-	-
41	-	1	-	-
44	1	-	-	-
<b>Total</b>	<b>1.948</b>	<b>4.441</b>	<b>6.251</b>	<b>6.561</b>

Fig. 1 Repeat Landscape do transposon En/Spm em *Apareiodon* sp. O gráfico mostra no eixo X a divergência para cada consenso em relação ao número de cópias no genoma, apresentada no eixo Y. a) macho b) fêmea.

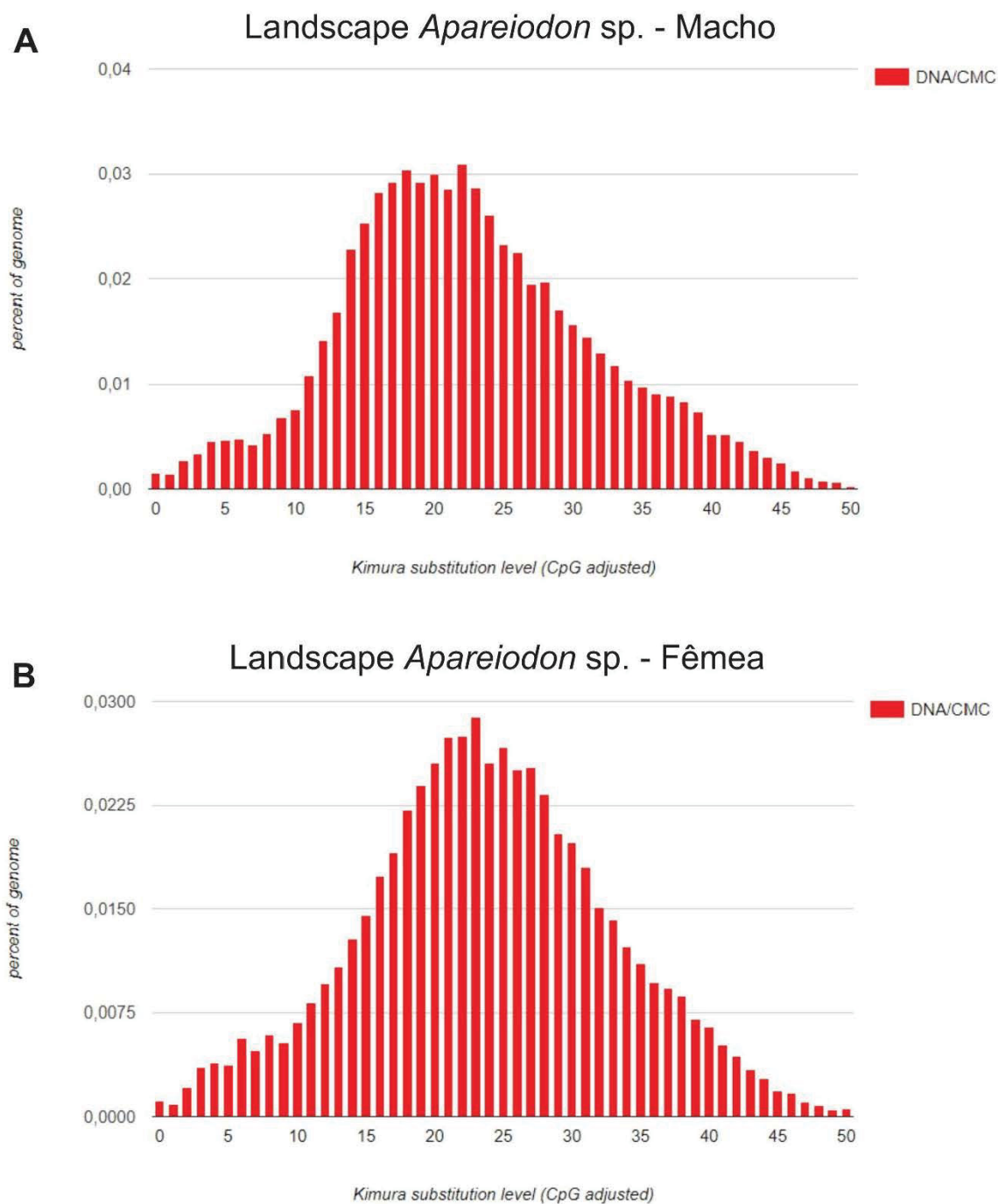




Fig. 2 Localização *in situ* do transposon En/Spm (vermelho) em cariótipos de fêmeas de espécies da família Parodontidae: Em (a) *A. piracicabae*, (b) *P. nasus*, (c) *Apareiodon* sp., (d) *P. hilarii* e, (e) *A. affinis*.

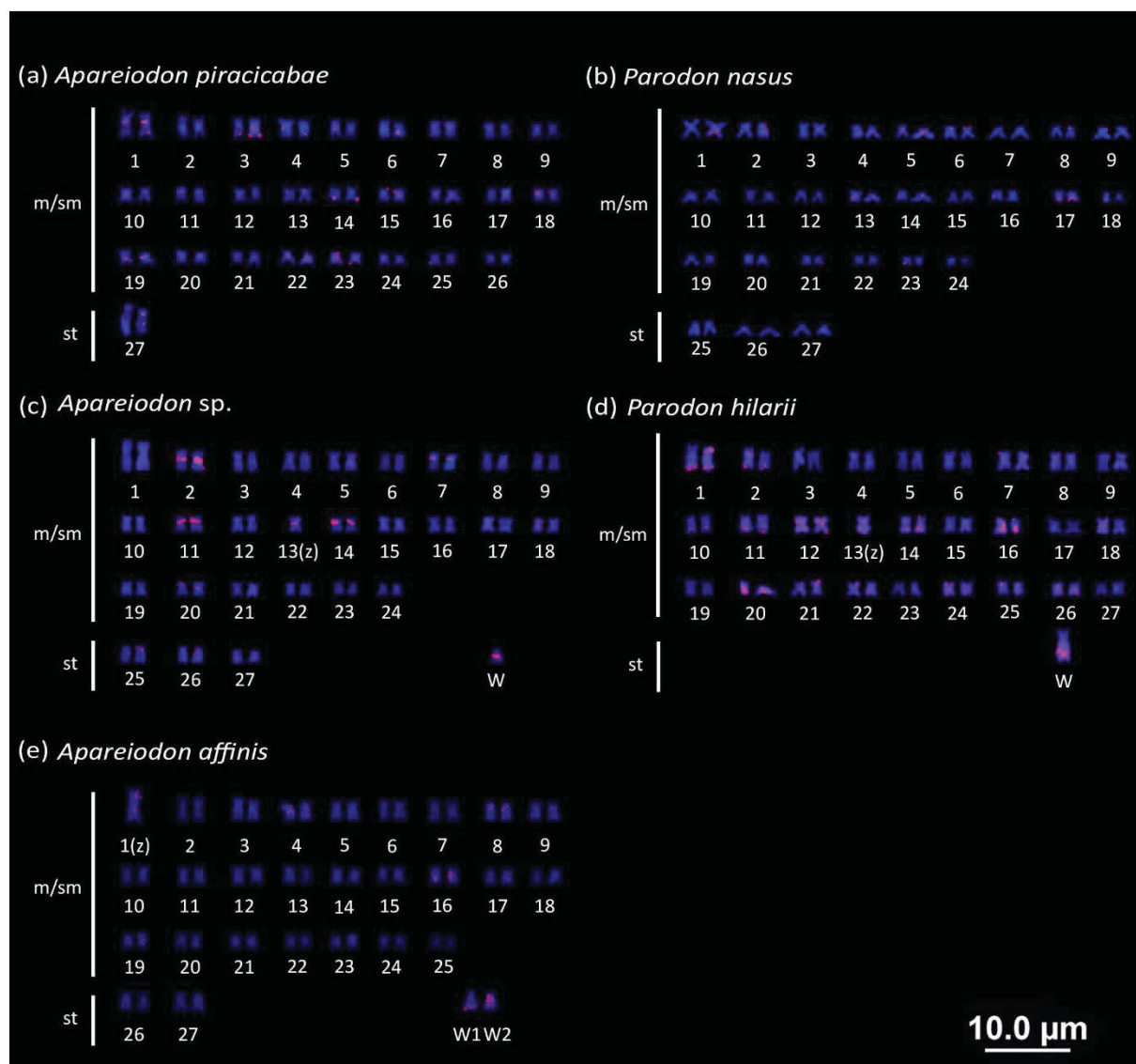


Fig. 3 Localização *in situ* dos microssatélites (GA)<sub>15</sub> e (TA)<sub>15</sub> (verde) em metáfases de fêmeas de espécies de Parodontidae: Em (a) *A. piracicabae*, (b) *P. nasus*, (c) *Apareiodon* sp., (d) *P. hilarii* e, (e) *A. affinis*.

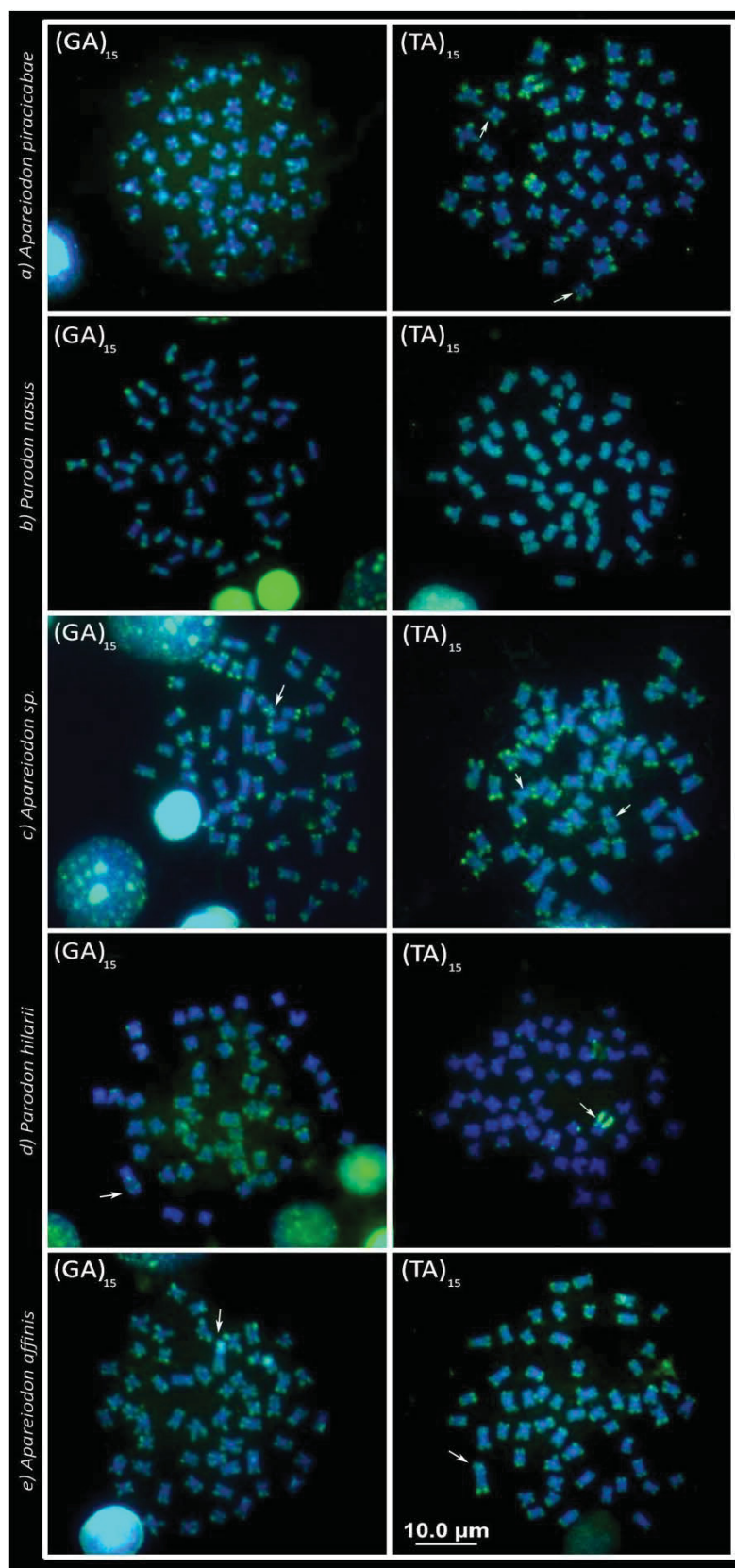


Tabela suplementar 1 Nome das espécies de Parodontidae, localidade de amostragem, distribuição em bacias hidrográficas e número de espécies estudadas.

<b>Espécies</b>	<b>Rio, Estado</b>	<b>Bacia Hidrográfica</b>	<b>Número de ♂</b>	<b>Número de ♀</b>	<b>GPS</b>
<i>A. piracicabae</i>	Mogi-Guaçu - SP	Alto Paraná	8	6	-21°55'38.6" e -47°21'58.4"
<i>P. nasus</i>	Mogi Guaçu - SP	Alto Paraná	11	9	-21°54'58" e -47°22'58"
<i>Apareiodon</i> sp.	Verde - PR	Alto Paraná	14	12	-25°04'35" e -50°04'03"
<i>P. hylarii</i>	São Francisco - MG	São Francisco	12	8	-17°21'17" e -44°57'15"
<i>A. affinis</i>	Passa Cinco - SP	Alto Paraná	10	7	-22°25'26" e -47°41'56"





Fig. Suplementar 1a TIRs do elemento transponível En/Spm no genoma macho de *Apareiodon* sp. similares às sequências (TIR-5' CACAGCAAAA.....AGTGT.AA.T....T e TIR3' GTGTCGTTTT.....CACAAATT.....). Análise dos 30 pares de bases das extremidades das sequências (destaque em vermelho). Numeração à esquerda refere-se à identificação de cada sequência no genoma. As regiões conservadas (5' CACAGCAAAA.....AGTGT.AA.T....T e 3' GTGTCGTTTT.....CACAAATT.....) estão destacadas em azul.

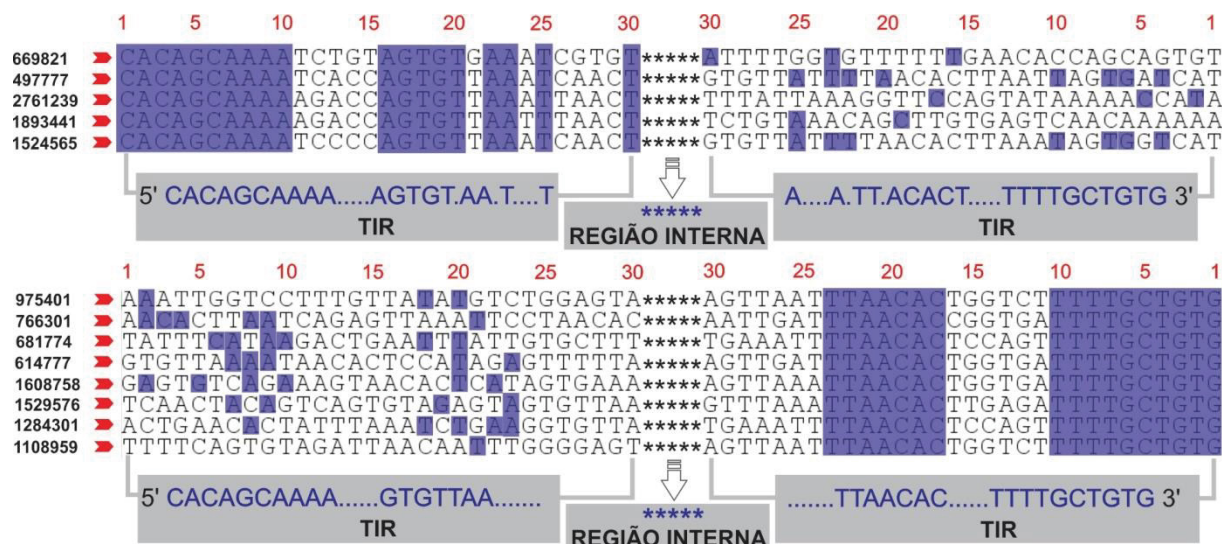




Fig. Suplementar 1b TIRs do elemento transponível En/Spm no genoma fêmea de *Apareiodon* sp. (TIR-5' CACAGCAAAA.....AGTGT.AA.T....T e TIR3' GTGTCGTTTT.....CACAATT.....). Análise dos 30 pares de bases das extremidades das sequências (destaque em vermelho). Numeração à esquerda refere-se à identificação de cada sequência no genoma. As regiões conservadas (5' CACAGCAAAA.....AGTGT.AA.T....T e 3' GTGTCGTTTT.....CACAATT.....) estão destacadas em azul.

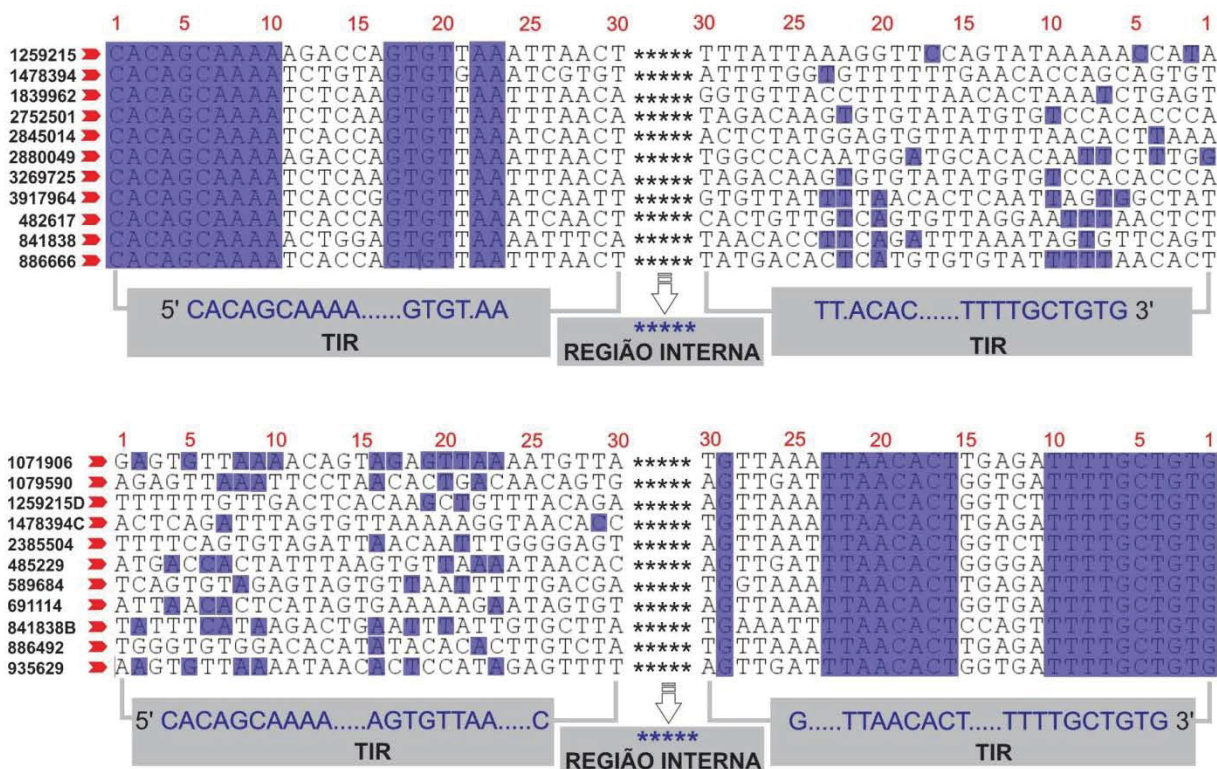




Fig. Suplementar 1c TIRs do elemento transponível En/Spm no genoma macho de *Aporeiodon* sp. similares às sequências (TIR-5' CACTGTAAAA e TIR-3' GTGACATTTT). Análise dos 30 pares de bases das extremidades das sequências (destaque em vermelho). Numeração à esquerda refere-se à identificação de cada sequência. As regiões conservadas (5' CACTGTAAAA e 3' GTGACATTTT) estão destacadas em azul.

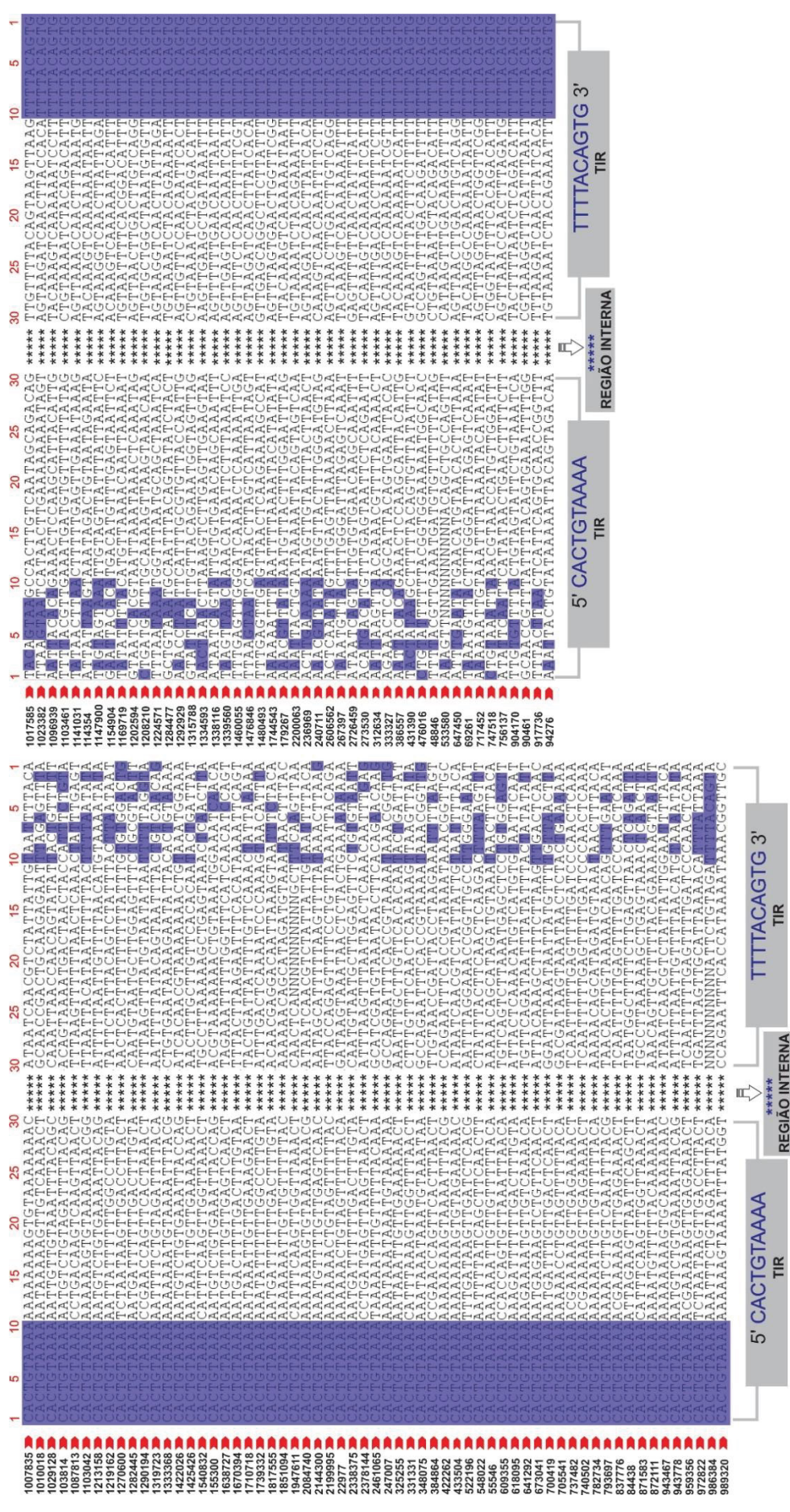
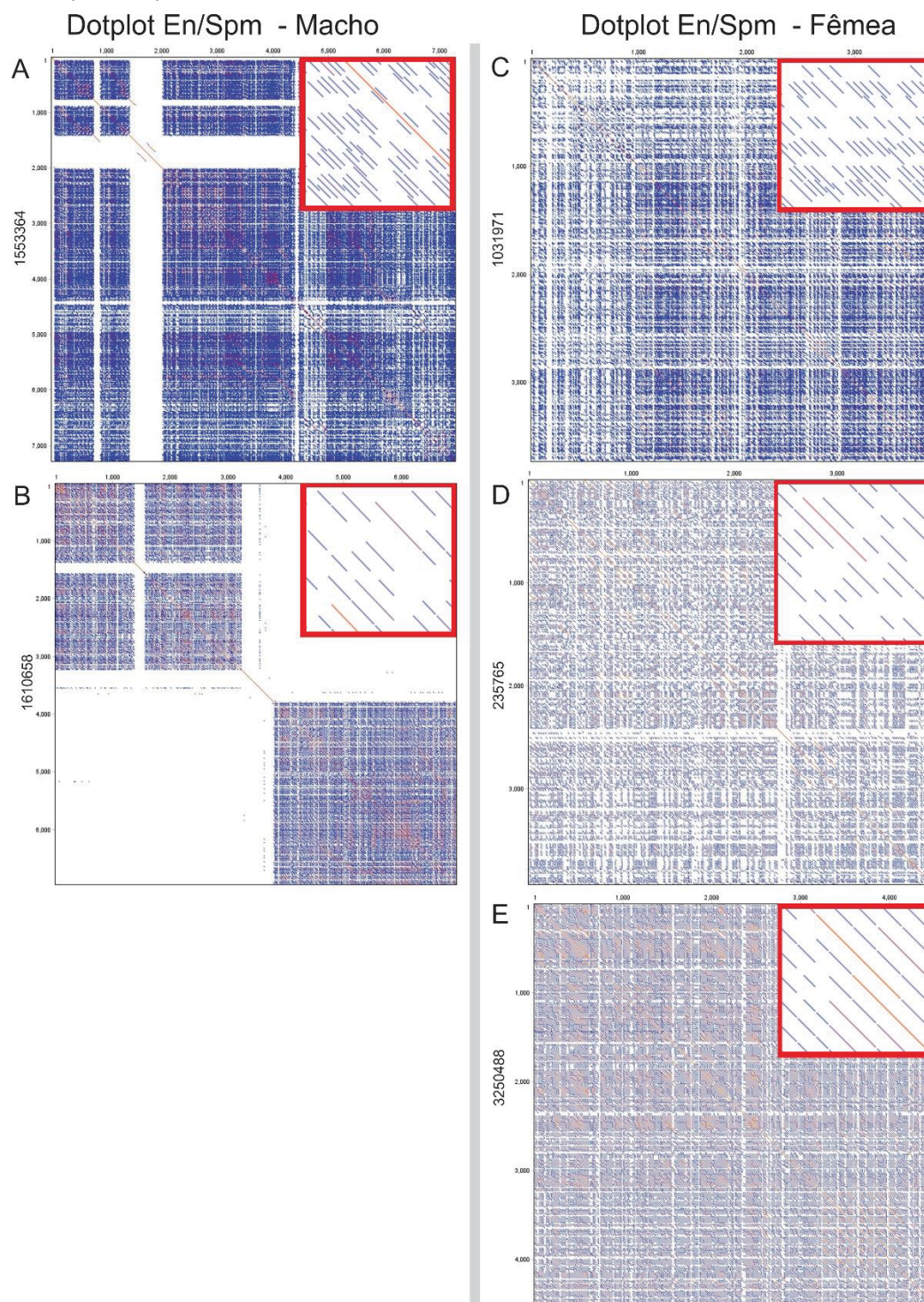








Fig. suplementar 2 Representação gráfica da análise de densidade de repetições *in tandem* em *contigs* contendo mais de 10 sequências En/Spm. Em (a, b) *contigs* 1553664 e 1610658 do genoma macho contendo alta densidade de repetições *in tandem*; em (c – e) *contigs* 1031971, 235765 e 3250488 do genoma fêmea contendo alta densidade de repetições *in tandem*. As linhas diagonais em azul representam fragmentos curtos (< que 100 pb) e as linhas diagonais em vermelho representam unidades > que 100 pb.



## 6 CONSIDERAÇÕES FINAIS

A constante busca por uma melhor compreensão da estrutura e funcionamento dos genomas demandou, nos últimos anos, um avanço significativo das novas tecnologias de sequenciamento e dos recursos computacionais a fim de decifrar de forma mais rápida e eficiente toda a complexidade que envolve os genomas eucarióticos. Neste sentido, a aplicação da tecnologia tem se tornado parte da rotina dos geneticistas, sendo essenciais para análises integradas e mais resolutivas, principalmente em estudos com DNAs repetitivos em genomas complexos.

Este estudo apresentou uma abordagem integrada de sequenciamento em larga escala, uso da bioinformática e dados citogenéticos a fim de contribuir para uma melhor compreensão do genoma de *Parodontidae*, principalmente no que se refere à fração repetitiva e origem do sistema de cromossomos sexuais heteromórficos do grupo.

Neste estudo, os elementos transponíveis Gypsy e En/Spm se apresentaram em diferentes níveis de degeneração molecular, com cópias defeituosas e sem capacidade de mobilidade. No entanto, evidências sugerem a participação dos TEs na gênese de expansões microssatélites, na origem de um DNA satélite e, na diferenciação de cromossomos sexuais. Em outra via, a análise de domínios proteicos derivados de TEs mostraram sequências que podem ter sido co-optadas para uma nova função do genoma de *Apareiodon* sp. Estes dados mostraram a complexidade da dinâmica dos DNAs repetitivos, abrindo novas perspectivas de trabalho na “mineração” destes dados genômicos. Este estudo acrescenta descobertas genômicas/moleculares para a diferenciação dos cromossomos sexuais em *Parodontidae* e apresenta uma primeira análise em perspectiva genômica para a elucidação do papel dos elementos repetitivos em peixes neotropicais.

## REFERÊNCIAS

- ALTINKUT, A.; RASKINA, O.; NEVO, E.; BELYAYEV, A. En/Spm-Like Transposons in Poaceae Species: Transposase Sequence Variability and Chromosomal Distribution. **Cellular and Molecular Biology Letters**, Wrocław, v. 11, n. 2, p. 214-230, jun. 2006.
- ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. **Journal of Molecular Biology**, London, v. 215, n. 3, p. 403-404, oct. 1990.
- AMOS, W. A comparative approach to study the evolution of microsatellites. In: GOLDSTEIN, D. B.; SCHLÖTTERER, C. (Ed.). **Microsatellites: Evolution and applications**. Oxford: Oxford University Press, 1999. p. 60-79.
- AMSELEM, J.; CORNUT, G.; CHOISNE, N.; ALAUX, M.; ALFAMA-DEPAUW, F.; JAMILLOUX, V.; MAUMUS, F.; LETELLIER, T.; LUYTEN, I.; POMMIER, C.; ADAM-BLONDON, A. F.; QUESNEVILLE, H. RepetDB: a unified resource for transposable element references. **Mobile DNA**, London, ano 10, v. 1, n. 6, p. 1-8, jan. 2019.
- APARICIO, S.; CHAPMAN, J.; STUPKA, E.; PUTNAM, N.; CHIA, J.M.; DEHAL, P.; CHRISTOFFELS, A.; RASH, S.; HOON, S.; SMIT, A.; GELPKE, M. D.; ROACH, J.; OH, T.; HO, I. Y.; WONG, M.; DETTER, C.; VERHOEF, F.; PREDKI, P.; TAY, A.; LUCAS, S.; RICHARDSON, P.; SMITH, S. F.; CLARK, M. S.; EDWARDS, Y. J.; DOGGETT, N.; ZHARKIKH, A.; TAVTIGIAN, S.V.; PRUSS, D.; BARNSTEAD, M.; EVANS, C.; BADEN H, POWELL J, GLUSMAN G, ROWEN L, HOOD L, TAN YH, ELGAR G, HAWKINS, T.; VENKATESH, B.; ROKHSAR, D.; BRENNER, S. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. **Science**, Washington, ano, 23, v. 297, n. 5585, p.1301-1310, aug. 2002.
- ARKHIPOVA, I. R. Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. **Mobile DNA**, London, v. 8, n. 19, p. 1-14, dec. 2017.
- BACHTROG, D.; WEISS, S.; ZANGERL, B.; BREM, G.; SCHLOTTERER, C. Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. **Molecular Biology and Evolution**, Oxford, v. 16, n. 5, p. 602-610, may. 1999.
- BAO, Z.; EDDY, S. R. Automated de novo identification of repeat sequence families in sequenced genomes. **Genome Research**. New York, v. 12, n. 8, p. 1269-1276, aug. 2002.
- BARBOSA, P.; OLIVEIRA, L. A.; PUCCI, M. B.; SANTOS, M. H.; MOREIRA-FILHO, O.; VICARI, M. R.; NOGAROTO, V.; ALMEIDA, M. C. de, ARTONI, R. F. Identification and chromosome mapping of repetitive elements in the *Astyanax scabripinnis* (Teleostei: Characidae) species complex. **Genetica**, 's-Gravenhage, v. 143, n. 1, p.55-62, feb. 2015.



- BARGHINI, E.; NATALI, L.; GIORDANI, T.; COSSU, R.M.; SCALABRIN, S.; CATTONARO, F.; ŠIMKOVÁ, H.; VRÁNA, J.; DOLEŽEL, J.; MORGANTE, M., CAVALLINI, A. LTR retrotransposon dynamics in the evolution of the olive (*Olea europaea*) genome. **DNA Research**, Tokyo, v. 22, n. 1, p. 91-100, nov. 2014.
- BEJERANO, G.; LOWE, C. B.; AHITUV, N.; KING, B.; SIEPEL, A.; SALAMA, S. R.; RUBIN, E. M.; KENT, W.J.; HAUSSLER, D. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. **Nature**, London, ano 4, v. 441, n. 7089, p. 87-90, may. 2006.
- BELL, G. I.; JURKA, J. The length distribution of perfect dimer repetitive DNA is consistent with its evolution by an unbiased single-step mutation process. **Journal of Molecular Evolution**. New York, v. 44, n. 4, p. 414-421, apr. 1997.
- BELLAFRONTE, E.; MARGARIDO, V. P.; MOREIRA-FILHO, O. Cytotaxonomy of *Parodon nasus* and *Parodon tortuosus* (Pisces, Characiformes). A case of synonymy confirmed by cytogenetic analyses. **Genetics and Molecular Biology**, São Paulo, v. 28, n. 4, p. 710-716, oct./dec. 2005.
- BELLAFRONTE, E.; SCHEMBERGER, M. O.; ARTONI, R. F.; MOREIRA-FILHO, M.; VICARI, M. R. Sex chromosome system ZZ/ZW in *Apareiodon hasemani* Eigenmann, 1916 (Characiformes, Parodontidae) and a derived chromosomal region. **Genetics and Molecular Biology**, São Paulo, v. 35, n.4, p. 770-776, dec. 2012.
- BELLAFRONTE, E.; SCHEMBERGER, M. O.; MOREIRA-FILHO, O.; ALMEIDA, M. C.; ARTONI, R. F.; MARGARIDO, V. P.; VICARI, M. R. Chromosomal markers in Parodontidae: an analysis of new and reviewed data with phylogenetic inferences. **Fish Biology and Fisheries**, Berlin, v. 21, n. 3, p. 559-570, sep. 2011.
- BELLAFRONTE, E.; VICARI, M. R.; ARTONI, R. F.; MARGARIDO, V. P.; MOREIRA-FILHO, O. Differentiated ZZ/ZW sex chromosomes in *Apareiodon ibitiensis* (Teleostei, Parodontidae): cytotaxonomy and biogeography. **Journal of Fish Biology**, Liverpool, v. 75, n. 9, p. 2313-2325, dec. 2009.
- BENACHENHOU, F.; BLIKSTAD, V.; BLOMBERG, J. The phylogeny of orthoretroviral long terminal repeats (LTRs). **Gene**, Amsterdam, ano 15, v. 448, n. 2, p. 134-138, dec. 2009.
- BENACHENHOU, F.; SPERBER, G. O.; BONGCAM-RUDLOFF, E.; ANDERSSON G.; BOEKE, J. D.; BLOMBERG, J. Conserved structure and inferred evolutionary history of long terminal repeats (LTRs). **Mobile DNA**, London, v. 4, n. 3, p. 1-15, feb. 2013.
- BERGMAN, C. M.; QUESNEVILLE, H. Discovering and detecting transposable elements in genome sequences. **Briefings in bioinformatics**, London, v. 8, n. 6, p. 382-392, nov. 2007.
- BERTOLLO, L. A. C.; CIOFFI, M. B.; MOREIRA-FILHO, O. Direct Chromosome preparations from freshwater teleost fishes. In: OZOUF-COSTAZ, C.; PISANO, E.;

- FORESTI, F., TOLEDO, L. F. de A. (Eds). **Fish techniques, Ray-Fin Fishes and Chondrichthyans**. New York: CRC Press, 2015. p. 21-26.
- BIÉMONT, C. A Brief History of the Status of Transposable Elements: from Junk DNA to Major Players in Evolution. **Genetics**, Austin, v. 186, n. 4, p. 1085-1093, dec. 2010.
- BIÉMONT, C.; VIEIRA, C.; HOOGLAND, C.; CIZERON, G.; LOEVENBRUCK, C.; ARNAULT, C.; CARANTE, J. P. Maintenance of transposable element copy number in natural populations of *Drosophila melanogaster* and *D. simulans*. **Genetica**, s'-Gravenhage, v. 100, n. 1-3, p. 161-166, feb. 1997.
- BLAUTH, M. L.; BRUNO, R.V.; ABDLHAY, E.; LORETO, E. L. S.; VALENTE, L. S. Detection of P element transcripts in embryos of *Drosophila melanogaster* e *D. willistoni*. **Annals of the Brazilian Academy of Sciences**, Rio de Janeiro, v. 81, n. 4, p. 679-689, dec. 2009.
- BOER, J. G.; YAZAWA, R.; DAVIDSON, W. S.; KOOP, B. F. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. **BioMed Central Genomics**, London, v. 8, n. 422, p. 1-10, nov. 2007.
- BRITTEN, R. J.; KOHNE, D. E. Nucleotide sequence repetition in DNA. **Carnegie Institution Wash Year Book**, Washington, v. 65, p. 78-106, feb. 1966.
- BRITTEN, R. J.; KOHNE, D. E. Repeat sequences in DNA. **Science**, Washington, v. 161, n. 3481, p. 529-540, aug. 1968.
- BUSCHIAZZO, E.; GEMMELL, N. J. The rise, fall and renaissance of microsatellites in eukaryotic genomes. **Bioessays**, Cambridge, v. 28, n. 10, p. 1040-1050, oct. 2006.
- BUSHMAN, F. D.; ENGELMAN, A.; PALMER, I.; WINGFIELD, P.; CRAIGIE, R. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. **Proceedings of the National Academy of Sciences of the United States of America**. Washington, v. 90, n. 8, p. 3428-3432, apr. 1993.
- BUSSIEK, M.; HOISCHEN, C.; DIEKMANN, S.; BENNINK, M. L. Sequence-specific physical properties of African green monkey alpha-satellite DNA contribute to centromeric heterochromatin formation. **Journal of Structural Biology**, Amsterdam, v. 167, n. 1, p. 36-46, jul. 2009.
- CÁCERES, M.; PUIG, M.; RUIZ, A. Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. **Genome Research**, New York, v. 11, n. 8, p. 1353-1364, sep. 2001.
- CÁCERES, M.; RANZ, J. M.; BARBADILLA, A.; LONG, M.; RUIZ, A. Generation of a widespread *Drosophila* inversion by a transposable element. **Science**, Washington, ano 16, v. 285, n. 5426, p. 415-418, jul. 1999.



- CAI, M. L.; ZHENG, R.L.; CAFFREY, M.; CRAIGIE, R.; CLORE, G.M.; GRONENBORN, A. M. Solution structure of the N-terminal zincbinding domain of HIV-1 integrase. **Nature structural & molecular biology**, New York, v. 4, n. 7, p. 567-577, jul. 1997.
- CAO, L.; YIN, G.; CAO, Z.; BING, X.; DING, W. Identification and characterization of a LTR retrotransposon from the genome of *Cyprinus carpio* var. Jian. **Genetica**, s'-Gravenhage, v. 144, n.3, p. 325-333, jun. 2016.
- CAPY, P.; BAZIN, C.; HIGUET, D.; LANGIN, T. **Dynamics and evolution of transposable elements**. Ann Arbor: Landes Bioscience, 1998. p.197.
- CASALS, F.; CÁCERES, M.; RUIZ, A. The foldback-like transposon Galileo is involved in the generation of two different natural chromosomal inversions of *Drosophila buzzatii*. **Molecular Biology and Evolution**, Oxford, v. 20, n. 5, p. 674-685, may. 2003.
- CASTRO, J. P.; CARARETO, C. M. *Drosophila melanogaster* P transposable elements: mechanisms of transposition and regulation. **Genetica**, s'-Gravenhage, v. 121, n. 12, p.107-118, jun. 2004.
- CAVALLI, G.; PARO, R. Chromo-Domain Proteins: Linking Chromatin Structure to Epigenetic Regulation. **Current Opinion in Cell Biology**, London, v. 10, n. 3, p. 354-60, jun. 1998.
- CAVALLINI, M. M.; BERTOLLO, L. A. C. Indução de mitoses em *Hoplias* cf. *malabaricus* (Teleostei, Characiformes, Erythrinidae). In: Simpósio de Citogenética Evolutiva e Aplicada de Peixes Neotropicais, II, 1988, Universidade Estadual de Maringá. **Anais do Simpósio de Citogenética Evolutiva e Aplicada de Peixes Neotropicais**. Maringá, 1988. p. 32.
- CENTOFANTE, L.; BERTOLLO, L. A. C.; MOREIRA-FILHO, O. ZZ/ZW sex chromosome system in new species of the genus *Parodon* (Pisces, Parodontidae). **Caryologia**, Firenze, v. 54, p. 139-150, feb. 2002.
- CHALOPIN, D.; NAVILLE, M.; PLARD, F.; GALIANA, D.; VOLFF, J. N. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. **Genome Biology Evolution**, Oxford, ano 9, v. 7, n. 2, p. 567-580, jan. 2015.
- CHIU, T. K.; DAVIES, D. R. Structure and Function of HIV-1 Integrase". **Current Topics in Medicinal Chemistry**, Xarja, v. 4, n. 9, p. 965-977, may. 2004.
- CHUONG, E. B.; ELDE, N. C.; FESCHOTTE, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. **Science**, New York, ano 351, v. 4, n. 6277, p. 1083-1087, mar. 2016.
- COAN, R.; MARTINS, C. Structural and functional analysis of transposable elements focusing on B chromosomes of the cichlid fish *Astatotilapia latifasciata*. **Genes**, Basel, v. 9, p. 269-287, may. 2018.

- COATES, B. S.; SUMERFORD, D. V.; HELLMICH, R. L.; LEWIS, L. C. Helitron-like transposon superfamily from *Lepidoptera disrupts* (GAAA)<sub>n</sub> microsatellites and is responsible for flanking sequence similarity within a microsatellite family. **Journal of molecular evolution**. Berlin, v. 70, n. 3, p. 275-288, mar. 2010.
- COCCA, E.; IORIO, S. de; CAPRIGLIONE, T. Identification of a novel helitron transposon in the genome of Antarctic fish. **Molecular Phylogenetics and Evolution**, Cambridge v. 58, n. 3, p. 439-446, jan. 2011.
- CRAIG, N. L. Mobile DNA: an introduction. In: CRAIG, N.; CRAIGIE, R.; GELLERT, M.; LAMBOWITZ, A. (Ed.). **Mobile DNA II**. Washington: ASM Press, 2002. p. 3-11.
- CUYPERS, H.; DASH, S.; PETERSON, P.; SAEDLER, H.; GIERL, A. The defective En -1102 element encodes a product reducing the mutability of the En/Spm transposable element of *Zea mays*. **The EMBO Journal**, Oxford, v. 7, n. 10, p. 2953-2960, oct. 1988.
- DALMAU, J.; GULTEKIN, S. H.; VOLTZ, R.; HOARD, R.; DESCHAMPS, T.; BALMACEDA, C.; BATCHELOR, T.; GERSTNER, E.; EICHEN, J.; FRENNIER, J.; POSNER, J. B.; ROSENFELD, M. R. Ma1, a novel neuron- and testis-specific protein, is recognized by the serum of patients with paraneoplastic neurological disorders, **Brain**, London, v. 122, n. 1, p. 27-39, jan. 1999.
- DE LA HERRÁN, R.; FONTANA, F.; LANFREDI, M.; CONGIU, L.; LEIS, M.; ROSSI, R.; RUIZ REJÓN, C.; RUIZ REJÓN, M.; GARRIDO-RAMOS, M. A. Slow rates of evolution and sequence homogenization in an ancient satellite DNA family of sturgeons. **Molecular Biology of Evolution**, Chicago, v. 18, n. 3, p. 432-436, mar. 2001.
- DOOLITTLE, R. F.; FENG, D. F.; JOHNSON, M. S.; MCCLURE, M. A. Origins and evolutionary relationships of retroviruses. **The Quarterly Review of Biology**, Chicago, v. 64, n. 1, p. 1-30, mar. 1989.
- DORER, D. R.; HENIKOFF, S. Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. **Cell Press**, Cambridge, v. 77, n. 7, p. 993-1002, jul. 1994.
- DOVER, G. A. Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. **Trends in Genetics**, Cambridge, v. 2, n. 6, p. 159-165, dec. 1986.
- DOYLE, J. J.; DOYLE, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **Phytochemistry Bulletin**, Madison, v. 19, p.11-15, feb. 1990.
- DRUMMOND, A. J.; ASHTON, B.; BUXTON, S.; CHEUNG, M.; COOPER, A.; HELED, J.; KEARSE, M.; MOIR, R.; STONES-HAVAS, S.; STURROCK, S.; THIERER, T.; WILSON, A. **Geneious**, versão 4.8.5., 2009.

- DUVERNELL, D. D.; PRYOR, S. R.; ADAMS, S. M. Teleost fish genomes contain a diverse array of L1 retrotransposon lineages that exhibit a low copy number and high rate of turnover. **Journal of Molecular Evolution**, Berlin, v. 59, n. 3, p. 298-308, sep. 2004.
- EDELSTEIN, L. C, COLLINS, T. The SCAN domain family of zinc finger transcription factors. **Gene**, Amsterdam, v. 359, p. 1-17, sep. 2005.
- EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, London, v. 32, n. 5, p. 1792-1797, mar. 2004.
- EICHLER, E. E.; SANKOFF, D. Structural dynamics of eukaryotic chromosome evolution. **Science**, New York, ano 8, v. 301, n. 5634, p. 793-797, aug. 2003.
- EICKBUSH, T. H.; JAMBURUTHUGODA, V. K. The diversity of retrotransposons and the properties of their reverse transcriptases. **Virus**, Amsterdam, v. 134, n. 1-2, p. 221-231, jun. 2008.
- EICKBUSH, T. H.; MALIK, H. S. Evolution of retrotransposons. In: CRAIG, N.; CRAIGIE, R.; GELLERT, M.; LAMBOWITZ, A. (Ed.). **Mobile DNA II**. Washington: ASM Press, 2002. p. 1111-1144.
- EIJKELENBOOM, A.; SPRANGERS, R.; HARD, K.; LUTZKE, R. A. P.; PLASTERK, R. H. A.; BOELEN, R.; KAPTEIN, R. Refined solution structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase. **Proteins, structure, function, and genetics**, New York, ano 1, v. 36, n 4, p. 556-564, sep. 1999.
- EL-GEHALI, S.; MISTRY, J.; BATEMAN, A.; EDDY, S. R.; LUCIANI, A.; POTTER, S. C.; QURESHI, M.; RICHARDSON, L. J.; SALAZAR, G. A.; SMART, A.; SONNHAMMER, E. L. L.; HIRSH, L.; PALADIN, L.; PIOVESAN, D.; TOSATTO, S. C. E.; FINN, R. D. The Pfam protein families database in 2019. **Nucleic Acids Research**, London, ano 8, v. 47, n. D1, p. D427-D432, jan. 2019.
- ELLISON, C. E.; BACHTROG, D. Dosage compensation via transposable element mediated rewiring of a regulatory network. **Science**, New York, ano 15, v. 342, n. 6160, p. 846-850, nov. 2013.
- EMERA, D.; WAGNER, G. P. Transformation of a transposon into a derived prolactin promoter with function during human pregnancy. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, ano 10, v. 109, n. 28, p. 11246-11251, jun. 2012.
- ESCHMEYER, W. N.; FONG, J. D. **Species of Fishes by family/subfamily**. Disponível em: <http://research.calacademy.org/redirect?url=http://researcharchive.calacademy.org/research/Ichthyology/catalog/SpeciesByFamily.asp>. Acesso em: 15 jan. 2019.

- EVGEN'EV, M. B.; YENIKOLOPOV, G. N.; PEUNOVA, N. I.; ILYIN, Y. V.  
Transposition of Mobile Genetic Elements in Interspecific Hybrids of *Drosophila*  
**Chromosoma**, Berlin, v. 85, n. 3, p. 375-386, jul.1982.
- FALCÃO, J. N.; MOREIRA-FILHO, O.; BERTOLLO, L. A. C. An additional  
chromosome in two fish species. **Brazilian Journal of Genetics**, São Paulo, v. 7,  
n. 1, p. 109-118, feb. 1984.
- FANTI, L.; DORER, D. R.; BERLOCO, M.; HENIKOFF, S.; PIMPINELLI, S.  
Heterochromatin protein 1 binds transgene arrays. **Chromosoma**, Berlin, v. 107,  
n. 5, p. 286-292, nov. 1998.
- FARRÉ, M.; BOSCH, M.; LÓPEZ-GIRÁLDEZ, F.; PONSÀ, M.; RUIZ-HERRERA, A.,  
Assessing the role of tandem repeats in shaping the genomic architecture of great  
apes. **Public Library of Science One**, San Francisco, v. 6, n. 11, p. e27239, nov.  
2011.
- FERNÁNDEZ-MEDINA, R. D.; RIBEIRO, J. M. C.; CARARETO, C. M. A.;  
VELASQUE, L.; STRUCHINER, C. J. Losing identity: structural diversity of  
transposable elements belonging to different classes in the genome of *Anopheles*  
*gambiae*. **BioMed Central Genomics**, London, v. 13, n. 1, p. 272-286, jun. 2012.
- FESCHOTTE, C. The contribution of transposable elements to the evolution of  
regulatory networks. **Nature Reviews Genetics**, London, v. 9, n. 5, p. 397-405,  
may. 2008.
- FESCHOTTE, C.; PRITHAM, E. J. DNA Transposons and the Evolution of Eukaryotic  
Genomes. **Annual Review of Genetics**, Arlington, v. 41, n. 1, p. 331-68, dec.  
2007.
- FESCHOTTE, C.; ZHANG, X.; WESSLER, S. R. Miniature Inverted-repeat  
Transposable Elements and their relationship to established DNA transposons. In:  
(CRAIG, N. L.; CRAIGIE, R.; GELLERT, M.; LAMBOWITZ, A. M.) **Mobile DNA II**.  
Washington: ASM Press, 2002. p. 1147-1158.
- FINNEGAN, D. J. Eukaryotic transposable elements and genome evolution. **Trends**  
**in Genetics**, Cambridge, v. 5, n. 4, p. 103-107, apr. 1989.
- FISCHER, C.; BOUNEAU, L.; COUTANCEAU, J.P.; WEISSENBACH, J.; OZOUF-  
COSTAZ, C.; VOLFF, J. N. Diversity and clustered distribution of  
retrotransposable elements in the compact genome of the pufferfish *Tetraodon*  
*nigroviridis*. **Cytogenetic and Genome Research**, Würzburg, v. 110, n 1-4, p.  
522-536, jul. 2005.
- FISCHER, C.; BOUNEAU, L.; COUTANCEAU, J. P.; WEISSENBACH, J.; VOLFF, J.  
N.; OZOUF-COSTAZ, C. Global heterochromatic colocalization of transposable  
elements with minisatellites in the compact genome of the pufferfish *Tetraodon*  
*nigroviridis*. **Gene**, Amsterdam, v. 336, n. 2, p. 175-183, jul. 2004.

- FRANK-KAMENETSKII, M. D.; MIRKIN, S. M. Triplex DNA structures. **Annual review of biochemistry**, Palo Alto, v. 64, n. 1, p. 65-95, jul. 1995.
- GAO, B.; SHEN, D.; XUE, S.; CHEN, C.; CUI, H.; SONG, C. The contribution of transposable elements to size variations between four teleost genomes. **Mobile DNA**, London, ano 9, v. 7, n. 4, p. 1-16, feb. 2016.
- GAO, D.; ABERNATHY, B.; ROHSAR, D.; SCHMUTZ, J.; JACKSON, S. A. Annotation and sequence diversity of transposable elements in common bean (*Phaseolus vulgaris*). **Frontiers in Plant Science**, Lausanne, v. 5, n. 339, p. 1-9, jul. 2014.
- GAO, X.; HAVECKER, E. R.; BARANOV, P. V.; ATKINS, J. F.; VOYTAS, D. F. Translational recoding signals between gag and pol in diverse LTR retrotransposons. **RNA**, New York, v. 9, n. 12, p. 1422-1430, dec. 2003.
- GARAVELLO, J. C. Descrição de *Apareiodon vittatus* sp. n. do Rio Iguaçu e comentários sobre as espécies do gênero *Apareiodon* Eigenmann, 1916 (Ostariophysi, Parodontidae). **Revista Brasileira de Biologia**, São Carlos, v. 37, n. 2, p. 447-455, feb. 1977.
- GBADEGESIN, M. A.; BEECHING, J. R. Enhancer/Suppressor mutator (En/Spm)-like transposable elements of cassava (*Manihot esculenta*) are transcriptionally inactive. **Genetics and Molecular Research**, Ribeirão Preto, v. 9, n. 2, p. 639-650, apr. 2010.
- GERLO, S.; DAVIS, J. R. E.; MAGER, D. L.; KOOIJMAN, R. Prolactin in man: a tale of two promoters. **Bioessays**, Cambridge, v. 28, n. 10, p. 1051-1055, oct. 2006.
- GERDES, P.; RICHARDSON, S. R.; MAGER, D. L.; FAULKNER, G. J. Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. **Genome biology**, London, ano, 9, v. 17, n. 100, p. 1-17, may. 2016.
- GIERL, A. The En/Spm transposable element of maize. In: SAEDLER, H.; GIERL, A. (Ed.) **Transposable Elements**. Current Topics in Microbiology and Immunology, v. 204, Berlin: Springer, 1996. p. 145-159.
- GIERL, A.; SCHWARZ-SOMMER, Z. S.; SAEDLER, H. Molecular interactions between the components of the En-1 transposable element system of *Zea mays*. **The EMBO journal**, Oxford, v. 4, n. 3, p. 579-583, mar. 1985.
- GLUGOSKI, L.; GIULIANO-CAETANO, L.; MOREIRA-FILHO, O.; VICARI, M. R.; NOGAROTO, V. Co-located hAT transposable element and 5S rDNA in an interstitial telomeric sequence suggest the formation of Robertsonian fusion in armored catfish. **Gene**, Amsterdam, v. 15, n. 650, p. 49-54, apr. 2018.
- GOODWIN, S.; MCPHERSON, J. D.; MCCOMBIE, W. R. Coming of age: ten years of nextgeneration sequencing Technologies. **Nature reviews Genetics**, London, ano 17, v. 17, n. 6, p. 333-351, may. 2016.

- GUO, X. M.; ZHANG, Q. Q.; SUN, Y. W.; JIANG, X. Y.; ZOU, S. M. Tc1-like Transposase Thm3 of Silver Carp (*Hypophthalmichthys molitrix*) Can Mediate Gene Transposition in the Genome of Blunt Snout Bream (*Megalobrama amblycephala*). **G3: genes - genomes - genetics**, Bethesda, ano 4, v. 5, n. 12, p. 2601-2610, oct. 2015.
- HAVECKER, E. R.; GAO, X.; VOYTAS, D. F. The diversity of LTR retrotransposons. **Genome Biology**, London, v. 5, n. 6, p. 225.1-225.6, may. 2004.
- HEIKKINEN, E.; LAUNONEN, V.; MÜLLER, E.; BACHMANN, L. The pvB370 BamHI satellite DNA family of the *Drosophila virilis* group and its evolutionary relation to mobile dispersed genetic pDv elements. **Journal of molecular evolution**, Berlin, v. 41, n. 5, p. 604-614, nov. 1995.
- HEREDIA, F. **Evolução do retroelemento Gypsy em espécies de Drosophyla e Zaprionus indianus: uma abordagem filogenética**. 214 f. Tese (Doutorado em Genética e Biologia Molecular). Setor de Ciências Biológicas, Universidade Federal Rio Grande do Sul, Porto Alegre, 2002.
- HERÉDIA, F.; LORETO, E. L. S.; VALENTE, V. L. S. Complex Evolution of gypsy in Drosophilid Species, **Molecular Biology and Evolution**, Oxford, v. 21, n. 10, p. 1831-1842, oct. 2004.
- HERNÁNDEZ-HERNÁNDEZ, A.; RINCÓN-ARANO, H.; RECILLAS-TARGA, F.; ORTIZ, R.; VALDES-QUEZADA, C.; ECHEVERRÍA, O. M.; BENAVENTE, R.; VÁZQUEZ-NIN, G. H. Differential distribution and association of repeat DNA sequences in the lateral element of the synaptonemal complex in rat spermatocytes. **Chromosoma**, Wien, v. 117, n. 1, p. 77-87, feb. 2008.
- HERPIN, A.; BRAASCH, I.; KRAEUSSLING, M.; SCHMIDT, C.; THOMA, E. C.; NAKAMURA, S.; TANAKA, M.; SCHAT, M. Transcriptional rewiring of the sex determining Dmrt1 gene duplicate by transposable elements. **Public Library of Science Genetics**, San Francisco, ano 12, v. 6, n. 2, p. p. e1000844, feb. 2010.
- HUA-VAN, A.; LE ROUZIC, A.; BOUTIN, T. S.; FILÉE, J.; CAPY, P. The struggle for life of the genome's selfish architects. **Biology Direct**, London, v. 6, n. 19, p. 1-2, mar. 2011.
- HUDA, A.; MARIÑO-RAMÍREZ, L.; JORDAN, K. Epigenetic histone modifications of human transposable elements: genome defense versus exaptation. **Mobile DNA**, London, v. 1, n. 2, p. 1-12, jan. 2010.
- HULL, R.; COVEY, S. N. Retroelements: propagation and adaptation. **Virus Genes**, New York, v. 11, n. 2-3, p. 105-118, jun. 1995.
- HURST, G. D., WERREN, J. H. The role of selfish genetic elements in eukaryotic evolution. **Nature reviews Genetics**, London, v. 2, n. 8, p. 597-606, aug. 2001.
- HURST, G. D. D.; SCHILTHUIZEN, M. Selfish genetic elements and speciation. **Heredity**, Basingstoke, v. 8, p. 2-8, jan. 1998.



- INGENITO, L. F. S. **Análise filogenética da família Parodontidae (Teleostei, Characiformes)**, 146 f. Tese (Doutorado em Zoologia). Setor de Ciências Biológicas, Museu Nacional. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.
- IWASAKI, S.; SUZUKI, S.; PELEKANOS, M.; CLARK, H.; ONO, R.; SHAW, G.; RENFREE, M. B.; KANEKO-ISHINO, T.; ISHINO, F. Identification of a novel PNMA-MS1 gene in marsupials suggests the LTR Retrotransposon-Derived PNMA genes evolved differently in marsupials and Eutherians. **DNA research**, Tokyo, v. 20, n. 5, p. 425-436, oct. 2003.
- IVICS, Z.; HACKETT, P. B.; PLASTERK, R. H.; IZSVA, Z. Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells. **Cell Press**, Cambridge, ano 14, v. 91, n. 4, p. 501-510, nov. 1997.
- JESUS, C. M.; BERTOLLO, L. A. C.; MOREIRA-FILHO, O. Comparative cytogenetics in *Apareiodon affinis* (Pisces, Characiformes) and considerations regarding diversification of the group. **Genetica**, s'-Gravenhage, v. 105, n. 1, p. 63-67, mar. 1999.
- JESUS, C. M.; MOREIRA-FILHO, O. Cytogenetic studies in some *Apareiodon* species (Pisces, Parodontidae). **Cytologia**, Tokyo, v. 65, n. 4, p. 397-402, sep. 2000a.
- JESUS, C. M.; MOREIRA-FILHO, O. Karyotypes of three species of Parodon (Teleostei, Parodontidae). **Ichthyological Exploration Freshwaters**, München, v. 11, p. 75-80, nov. 2000b.
- JIANG, N.; FERGUSON, A. A.; SLOTKIN, R. K.; LISCH, D. 2011 Pack-Mutator-like transposable elements (Pack-MULEs) induce directional modification of genes through biased insertion and DNA acquisition. **Proceedings of the Nacional Academy of Sciences**, Washington, v. 108, n. 4, p. 1537-1542, jan. 2011.
- JOHNSON, M. S.; MCCLURE, M. A.; FENG, D. F.; GRAY, J.; DOOLITTLE, R. F. Computer analysis of retroviral pal genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. **Proceedings of the Nacional Academy of Sciences**, Washington, v. 83, n. 20, p. 7648-7652, oct. 1986.
- JORGE, L. C.; MOREIRA-FILHO, O. Cytogenetic studies on *Apareiodon affinis* (Pisces, Characiformes) from Paraná river basin: Sex chromosomes and polymorphism. **Genetica**, s'-Gravenhage, v. 109, p. 267-273, oct. 2000.
- JORGE, L. C.; MOREIRA-FILHO, O. Nucleolar organizer regions as markers of chromosomal polymorphism in *Apareiodon affinis* (Pisces, Parodontidae). **Caryologia**, Firenze, v. 57, p. 203-207, dec. 2004.
- KAMBOL, R.; ABTHOLUDDIN, M. F. Genome structure and characterisation of an endogenous retrovirus from the zebrafish genome project database. In: ELLOUMI,

- M.; KÜNG, J.; LINIAL, M.; MURPHY, R. F.; SCHNEIDER, K.; TOMA, C. (Ed.). **Bioinformatics Research and Development**. Communications in Computer and Information Science, v.13, Berlin: Springer, 2008. p. 392-403.
- KANTEK, D. L.; CESTARI, M. M.; BERTOLLO, L. A.; MOREIRA-FILHO, O. M. Characterization of the constitutive heterochromatin of *Astyanax* sp. (Characiformes: Characidae) from the Upper Iguaçu River (PR). **Folia Biologica**, Warszawa, v. 57, n. 1-2, p. 37-41, feb. 2009.
- KAPITONOV, V. V.; JURKA, J. Molecular paleontology of transposable elements from *Arabidopsis thaliana*. **Genetics**, s'-Gravenhage, v. 107, n. 1-3, dec. 1999.
- KAPITONOV, V. V.; JURKA, J. A universal classification of eukaryotic transposable elements implemented in Repbase. **Nature Reviews Genetics**, London, v. 9, p. 411-412, may. 2008.
- KAPUSTA, A.; KRONENBERG, Z.; LYNCH, V. J.; ZHUO, X.; RAMSAY, L.; BOURQUE, G.; YANDELL, M.; FESCHOTTE, C. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. **Public Library of Science Genetics**, San Francisco, v. 9, n. 4, apr. 2013.
- KELKAR, Y. D.; ECKERT, K. A.; CHIAROMONTE, F.; MAKOVA, K. D. A matter of life or death: How microsatellites emerge in and vanish from the human genome. **Genome Research**, New York, v. 21, n. 12, p. 2038-2048, dec. 2011.
- KIDWELL, M. G. Transposable elements and the evolution of genome size in eukaryotes. **Genetica**, s'-Gravenhage, v. 115, n. 1, p. 49-63, may. 2002.
- KIDWELL, M. G.; HOLYOAKE, A. J. Transposon-induced hotspots for genomic instability. **Genome Research**, New York, v. 11, n. 8, p. 1321-1322, aug. 2001.
- KIDWELL, M. G.; KIDWELL, J. F.; SVED, J. A. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. **Genetics**, Austin, v. 86, n. 4, p. 813-833, aug. 1977.
- KIDWELL, M. G.; LISCH, D. Transposable elements as sources of variation in animals and plants. **Proceedings of the Nacional Academy of Sciences**, Washington, v. 94, n. 15, p. 7704-7711, jul. 1997.
- KIDWELL, M. G.; LISCH, D. R. Perspective: transposable elements, parasitic DNA, and genome evolution. **International Journal of organic Evolution**, Hoboken, v. 55, n. 1, p. 1-24, jan. 2001.
- KOLADE, O.; RAJI, A.; FAWOLE, I.; INGELBRECHT, I. Molecular Characterization of Type II Transposable Elements in Cowpea [*Vigna unguiculata* (L.) Walp]. **American Journal of Plant Sciences**, Wuhan, v. 6, n. 5, p. 767-776, apr. 2015.
- KOGA, A.; IIDA, A.; HORI, H.; SHIMADA, A.; SHIMA, A. Vertebrate DNA transposon as a natural mutator: the medaka fish Tol2 element contributes to genetic variation

without recognizable traces. **Molecular Biology and Evolution**, Oxford, v. 23, n. 7, p. 1414-1419, jul. 2006.

KOHANY, O.; GENTLES, A. J.; HANKUS, L.; JURKA, J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. **Bio MedCentral Bioinformatics**. London, v. 7, n. 474, p. 1-7, oct. 2006.

KOJIMA, K. K.; JURKA, J. Crypton transposons: identification of new diverse families and ancient domestication events. **Mobile DNA**, London, v. 2, n. 12, p. 1-17, oct. 2011.

KOMISSAROV, A. S.; GALKINA, S. A.; KOSHEL, E. I.; KULAK, M. M.; DYOMIN, G. A.; O'BRIEN, S. J.; GAGINSKAYA, E. R.; SAIFITDINOVA, A. F. New high copy tandem repeat in the content of the chicken W chromosome. **Chromosoma**, Berlin, v. 127, n. 1, p. 73-83, mar. 2018.

KOONIN, E. V.; ZHOU, S.; LUCCHESI, J. C. The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. **Nucleic Acids Research**, Oxford, v. 23, n. 21, p. 4229-4233, jan. 1995.

KRIMBAS, C. B.; POWELL, J. R. Introduction. In: KRIMBAS, C. B.; POWELL, J. R. (Ed.). **Drosophila Inversion Polymorphism**, Boca Raton: CRC Press, 1992. p. 1-52.

KURENOVA, E.; CHAMPION, L.; BIESSMANN, H.; MASON, J.M. Directional gene silencing induced by a complex subtelomeric satellite from Drosophila. **Chromosoma**, Wien, v. 107, n. 5, p. 311-20, nov. 1998.

LANDER, E. S.; LINTON, L. M.; BIRREN, B.; NUSBAUM, C.; ZODY, M. C.; BALDWIN, J.; DEVON, K.; DEWAR, K.; DOYLE, M.; FITZHUGH, W.; FUNKE, R.; GAGE, D.; HARRIS, K.; HEAFORD, A.; HOWLAND, J.; KANN, L.; LEHOCZKY, J.; LEVINE, R.; MCEWAN, P.; MCKERNAN, K.; MELDRIM, J.; MESIROV, J. P.; MIRANDA, C.; MORRIS, W.; NAYLOR, J.; RAYMOND, C.; ROSETTI, M.; SANTOS, R.; SHERIDAN, A.; SOUGNEZ, C.; STANGE-THOMANN, Y.; STOJANOVIC, N.; SUBRAMANIAN, A.; WYMAN, D.; ROGERS, J.; SULSTON, J.; AINSCOUGH, R.; BECK, S.; BENTLEY, D.; BURTON, J.; CLEE, C.; CARTER, N.; COULSON, A.; DEADMAN, R.; DELOUKAS, P.; DUNHAM, A.; DUNHAM, I.; DURBIN, R.; FRENCH, L.; GRAFHAM, D.; GREGORY, S.; HUBBARD, T.; HUMPHRAY, S.; HUNT, A.; JONES, M.; LLOYD, C.; MCMURRAY, A.; MATTHEWS, L.; MERCER, S.; MILNE, S.; MULLIKIN, J. C.; MUNGALL, A.; PLUMB, R.; ROSS, M.; SHOWNKEEN, R.; SIMS, S.; WATERSTON, R. H.; WILSON, R. K.; HILLIER, L. W.; MCPHERSON, J. D.; MARRA, M. A.; MARDIS, E. R.; FULTON, L. A.; CHINWALLA, A. T.; PEPIN, K. H.; GISH, W. R.; CHISSOE, S. L.; WENDL, M. C.; DELEHAUNTY, K. D.; MINER, T. L.; DELEHAUNTY, A.; KRAMER, J. B.; COOK, L. L.; FULTON, R. S.; JOHNSON, D. L.; MINX, P. J.; CLIFTON, S. W.; HAWKINS, T.; BRANSCOMB, E.; PREDKI, P.; RICHARDSON, P.; WENNING, S.; SLEZAK, T.; DOGGETT, N.; CHENG, J. F.; OLSEN, A.; LUCAS, S.; ELKIN, C.; UBERBACHER, E.; FRAZIER, M.; GIBBS, R. A.; MUZNY, D. M.; SCHERER, S. E.; BOUCK, J. B.; SODERGREN, E. J.; WORLEY, K. C.;

- RIVES, C. M.; GORRELL, J. H.; METZKER, M. L.; NAYLOR, S. L.; KUCHERLAPATI, R. S.; NELSON, D. L.; WEINSTOCK, G. M.; SAKAKI, Y.; FUJIYAMA, A.; HATTORI, M.; YADA, T.; TOYODA, A.; ITOH, T.; KAWAGOE, C.; WATANABE, H.; TOTOKI, Y.; TAYLOR, T.; WEISSENBAACH, J.; HEILIG, R.; SAURIN, W.; ARTIGUENAVE, F.; BROTTIER, P.; BRULS, T.; PELLETIER, E.; ROBERT, C.; WINCKER, P.; SMITH, D. R.; DOUCETTE-STAMM, L.; RUBENFIELD, M.; WEINSTOCK, K.; LEE, H. M.; DUBOIS, J.; ROSENTHAL, A.; PLATZER, M.; NYAKATURA, G.; TAUDIEN, S.; RUMP, A.; YANG, H.; YU, J.; WANG, J.; HUANG, G.; GU, J.; HOOD, L.; ROWEN, L.; MADAN, A.; QIN, S.; DAVIS, R. W.; FEDERSPIEL, N. A.; ABOLA, A. P.; PROCTOR, M. J.; MYERS, R. M.; SCHMUTZ, J.; DICKSON, M.; GRIMWOOD, J.; COX, D. R.; OLSON, M. V.; KAUL, R.; RAYMOND, C.; SHIMIZU, N.; KAWASAKI, K.; MINOSHIMA, S.; EVANS, G. A.; ATHANASIOU, M.; SCHULTZ, R.; ROE, B. A.; CHEN, F.; PAN, H.; RAMSER, J.; LEHRACH, H.; REINHARDT, R.; MCCOMBIE, W. R.; DE LA BASTIDE, M.; DEDHIA, N.; BLÖCKER, H.; HORNISCHER, K.; NORDSIEK, G.; AGARWALA, R.; ARAVIND, L.; BAILEY, J. A.; BATEMAN, A.; BATZOGLOU, S.; BIRNEY, E.; BORK, P.; BROWN, D. G.; BURGE, C. B.; CERUTTI, L.; CHEN, H. C.; CHURCH, D.; CLAMP, M.; COPLEY, R. R.; DOERKS, T.; EDDY, S. R.; EICHLER, E. E.; FUREY, T. S.; GALAGAN, J.; GILBERT, J. G.; HARMON, C.; HAYASHIZAKI, Y.; HAUSSLER, D.; HERMJAKOB, H.; HOKAMP, K.; JANG, W.; JOHNSON, L. S.; JONES, T. A.; KASIF, S.; KASPRYZK, A.; KENNEDY, S.; KENT, W. J.; KITTS, P.; KOONIN, E. V.; KORF, I.; KULP, D.; LANCET, D.; LOWE, T. M.; MCLYSAGHT, A.; MIKKELSEN, T.; MORAN, J. V.; MULDER, N.; POLLARA, V. J.; PONTING, C. P.; SCHULER, G.; SCHULTZ, J.; SLATER, G.; SMIT, A. F.; STUPKA, E.; SZUSTAKOWKI, J.; THIERRY-MIEG, D.; THIERRY-MIEG, J.; WAGNER, L.; WALLIS, J.; WHEELER, R.; WILLIAMS, A.; WOLF, Y. I.; WOLFE, K. H.; YANG, S. P.; YEH, R. F.; COLLINS, F.; GUYER, M. S.; PETERSON, J.; FELSENFELD, A.; WETTERSTRAND, K. A.; PATRINOS, A.; MORGAN, M. J.; DE JONG, P.; CATANESE, J. J.; OSOEGAWA, K.; SHIZUYA, H.; CHOI, S.; CHEN, Y. J.; SZUSTAKOWKI, J. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. **Nature**, London, v. 409, n. 6822, p. 860-921, feb. 2001.
- LANGLEY, C.; MONTGOMERY, E.; HUDSON, R.; KAPLAN, N.; CHARLESWORTH, B. On the role of unequal exchange in the containment of transposable element copy number. **Genetical Research**, London, v. 52, n. 3, p. 223-235, dec. 1988.
- LEVAN, A.; FREDGA, K.; SANDBERG, A. A. Nomenclature for centromeric position on chromosomes. **Hereditas**, London, v. 52, n. 2, p. 201-220, nov. 1964.
- LERAT E. Identifying repeats and transposable elements in sequence genomes: how to find your way through the dense forest of programs. **Heredity**, Basingstoke, v. 104, p. 520-533, nov. 2010.
- LEVINSON, G.; GUTMAN, G. A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. **Molecular Biology and Evolution**, Oxford, v. 4, n. 3, p. 203-221, may. 1987.
- LEWIN, B. **Gene XI**. 11<sup>th</sup> ed. Burlington: Jones & Bartlett Publishers, 2014.

- LI, S. F.; GUO, Y. J.; LI, J. R.; ZHANG, D. X.; WANG, B. X.; LI, N.; DENG, C. L.; GAO, W. J. The landscape of transposable elements and satellite DNAs in the genome of a dioecious plant spinach (*Spinacia oleracea* L.). **Mobile DNA**, London, v. 10, n. 1, p. 1-15, jan. 2019.
- LIEN, S.; KOOP, B. F.; SANDVE, S. R.; MILLER, J. R.; KENT, M. P.; NOME, T.; HVIDSTEN, T. F.; LEONG, J. S.; MINKLEY, D. R.; ZIMIN, A.; GRAMMES, F.; GROVE, H.; GJUVSLAND, A.; WALENZ, B.; HERMANSEN, R. A.; VON-SCHALBURG K.; RONDEAU, E. B.; DI-GENOVA, A.; SAMY, J. K. A.; VIK, J. O.; VIGELAND, M. D.; CALER, L.; GRIMHOLT, U.; JENTOFT, S.; VÅGE, D. I.; JONG, P. de.; MOEN, T.; BARANSKI, M.; PALT, Y.; SMITH, D. R.; YORKE, J. A.; NEDERBRAGT, A. J.; TOOMING-KLUNDERUD, A.; JAKOBSEN, K. S.; JIANG, X.; FAN, D.; HU, Y.; LIBERLES, D. A.; VIDAL, R.; ITURRA, P.; JONES, S. J.; JONASSEN, I.; MAASS, A.; OMHOLT, S. W.; DAVIDSON, W. S. The Atlantic Salmon Genome Provides Insights Into Rediploidization. **Nature**, London, v. 533, p. 200-205, may. 2016.
- LITT, M.; LUTY, J. A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. **American journal of human genetics**, Baltimore, v. 44, n. 3, p. 397-401, mar. 1989.
- LLORENS, C.; FUTAMI, R.; COVELLI, L.; DOMÍNGUEZ-ESCRIBÁ, L.; VIU, J. M.; TAMARIT, D.; AGUILAR-RODRÍGUEZ, J.; VICENTE-RIPOLLES, M.; FUSTER, G.; BERNET, G. P.; MAUMUS, F.; MUNOZ-POMER, A.; SEMPERE, J. M.; LATORRE, A.; MOYA, A. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. **Nucleic Acids Research**, London, v. 39, n. 1, jan. 2011.
- LODI, P. J.; ERNST, J. A.; KUSZEWSKI, J. HICKMAN, A. B.; ENGELMAN, A.; CRAIGIE, R.; CLORE, G. M.; GRONENBORN, A. M.; Solution Structure of the DNA-Binding Domain of Hiv-1 Integrase. **Biochemistry**, Washington, ano 8, v. 34, n. 31, p. 9826-9833, aug. 1995.
- LONGO, M. S.; CARONE, D. M.; NISC COMPARATIVE SEQUENCING PROGRAM; GREEN, E. D.; O'NEILL, M. J.; O'NEILL, R. J. Distinct retroelement classes define evolutionary breakpoints demarcating sites of evolutionary novelty. **BioMed Central Genomics**, London, v. 10, n. 334, p. 1-14, jul. 2009.
- LÓPEZ-FLORES, I.; GARRIDO-RAMOS, M. A. The repetitive DNA content of eukaryotic genomes. In: GARRIDO-RAMOS, M. (Ed.). **Repetitive DNA**. Genome Dynamics, v. 7, Basel: Karger, 2012. p. 1-28.
- LOWE, C. B.; HAUSSLER, D. 29 mammalian genomes reveal novel exaptations of mobile elements for likely regulatory functions in the human genome. **Public Library of Science One**, San Francisco, v. 7, n. 8, p. 1-31, aug. 2012.
- MACAS, J.; KEJNOVSKY, E. NEUMANN, P.; NOVAK, P.; KOBLIZKOVA, A.; VYSKOT, B. Next generation sequencing-based analysis of repetitive DNA in the model dioecious plant *Silene latifolia*. **Public Library of Science One**, San Francisco, v. 6, n. 11, p. 1-12, nov. 2011.



- MACAS, J.; NEUMANN, P.; NOVÁK, P.; JIANG, J. Global sequence characterization of rice centromeric satellite based on oligomer frequency analysis in largescale sequencing data. **Bioinformatics**, Oxford, v. 26, n. 17, p. 2101-2108, sep. 2010.
- MAGER, D. L.; STOYE, J. P. Mammalian endogenous retroviruses. **Microbiology spectrum**, Washington, v. 3, n. 1, p. 1-20, feb. 2015.
- MAIZEL, J. V.; LENK, R. P. Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proceedings of the National Academy of Sciences. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 78, n. 12, p. 7665-7669, dec. 1981.
- MALIK, H. S.; BURKE, W. D. EICKBUSH, T. H. The age and evolution of non-LTR retrotransposable elements. **Molecular biology and evolution**, Chicago, v. 16, n. 6, p. 793-805, jun. 1999.
- MARBURGER, S.; ALEXANDROU, M. A.; TAGGART, J. B.; CREER, S.; CARVALHO, G.; OLIVEIRA, C.; TAYLOR, M. I. Whole genome duplication and transposable element proliferation drive genome expansion in *Corydoradinae catfishes*. Proceedings, **Biological sciences / The Royal Society**. London, v. 285, n. 1872, p. 1-10, feb. 2018.
- MARDIS, E. R. Next-Generation DNA Sequencing Methods. **Annual Review of Genomics and Human Genetics**, Palo Alto, v. 9, p. 387-402, sep. 2008.
- MARTINS, C. Chromosomes and repetitive DNAs: a contribution to the knowledge of fish genome. In: PISANO, E.; OZOUF-COSTAZ, C.; FORESTI, F.; KAPOOR, B.G. (Ed.). **Fish Cytogenetics**. Enfield: Science Publisher, 2007. p. 421-453.
- MAZZUCHELLI, J.; MARTINS, C. Genomic organization of repetitive DNAs in the cichlid fish *Astronotus ocellatus*. **Genetica**, Dordrecht, v. 136, n. 3, p. 461-469, jul. 2009.
- MCCLINTOCK, B. The Origin and Behavior of Mutable Loci in Maize. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 36, n. 6, p. 344-355, jun. 1950.
- MCCLINTOCK, B. Mutations in Maize and Chromosomal Aberrations in Neurospora. **Annual Report of the Director of the Department of Genetics**, Washington, n. 53, p. 254-260, dec. 1954.
- MEDSTRAND, P.; VAN DE LAGEMAAT, L. N.; DUNN, C. A.; LANDRY, J. R.; SVENBACK, D.; MAGER, D. L. Impact of transposable elements on the evolution of mammalian gene regulation. **Cytogenetic and genome research**, Basel, v. 110, n. 1-4, p. 342-352, jul. 2005.
- MILLER, W. J.; MCDONALD, J. F.; PINSKER, W. Molecular domestication of mobile elements. In: Capy P. (Ed.). **Evolution and Impact of Transposable Elements**. Contemporary Issues in Genetics and Evolution, v. 6, Dordrecht: Springer, 1997. p. 261-270.



- MONOD, C.; AULNER, N.; CUVIER, O.; KÄS, E. Modification of position-effect variegation by competition for binding to Drosophila satellites. **European Molecular Biology Organization reports**, Oxford, v. 3, n. 8, p. 747-52, aug. 2002.
- MOREIRA-FILHO, O.; BERTOLLO, L. A. C.; GALETTI JUNIOR, P. M. Evidences for a multiple sex chromosome system with female heterogamety in *Apareiodon affinis* (Pisces, Parodontidae). **Caryologia**, Firenze, v. 33, p. 83-91, jan. 1980.
- MOREIRA-FILHO, O.; BERTOLLO, L. A. C.; GALETTI JUNIOR, P. M. Structure and variability of nucleolar organizer regions in Parodontidae fish. **Canadian Journal of Genetics and Cytology**, Ottawa, v. 26, n. 5, p. 564-568, oct. 1984.
- MOREIRA-FILHO, O.; BERTOLLO, L. A. C.; GALETTI JUNIOR, P. M. Karyotypic study of some species of family Parodontidae (Pisces-Cypriniformes). **Caryologia**, Firenze, v. 38, n. 1, p. 47-55, jul. 1985.
- MOREIRA-FILHO, O.; BERTOLLO, L. A. C.; GALETTI JUNIOR, P. M. Distribution of sex chromosome mechanisms in Neotropical fish and description of a ZZ/ZW system in *Parodon hilarii* (Parodontidae). **Caryologia**, Firenze, v. 46, p. 115-125, jan. 1993.
- MORGANTE, M.; HANAFEY, M.; POWELL, W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. **Nature Genetics**, New York, v. 30, n. 2, p. 194-200, feb. 2002.
- MORITA, E.; SUNDQUIST, W. I. Retrovirus Budding. **Annual review of cell and developmental biology**, Palo Alto, v. 20, p. 395-425, nov. 2004
- NADIR, E.; MARGALIT, H.; GALLILY, T.; BEN-SASSON, S. A. Microsatellite spreading in the human genome : evolutionary mechanisms and structural implications. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 93, n. 13, p. 6470-6475, jun. 1996.
- NASCIMENTO, V. D. do; COELHO, K. A.; NOGAROTO, V.; DE ALMEIDA, R. B.; ZIEMNICZAK, K.; CENTOFANTE, L.; PAVANELLI, C. S.; TORRES, R. A.; MOREIRA-FILHO, O.; VICARI, M. R. Do multiple karyomorphs and population genetics of freshwater darter characines (*Apareiodon affinis*) indicate chromosomal speciation?. **Zoologischer Anzeiger**, Berlin, v. 272, p. 93-103, jan. 2018.
- NASSIF, N.; PENNEY, J.; PAL, S.; ENGELS, W.; GLOOR, G. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. **Molecular and cellular biology**, Washington, v. 14, n. 3, p. 1613-1625, mar. 1994.
- NOMURA, H. Caracteres Merísticos e dados biológicos sobre o canivete *Parodon tortuosus* Eigenmann & Norris, 1990, do Rio Mogi Guaçu, São Paulo (Osteichthyes, Parodontidae). **Revista Brasileira de Biologia**, Rio de Janeiro, v. 39, p. 451-456, feb. 1979.

- NOORALI, S.; ROTAR, I. C.; LEWIS, C.; PESTANER, J. P.; PACE, D. G.; SISON, A.; BAGASRA, O. Role of HERV-W syncytin-1 in placentation and maintenance of human pregnancy. **Applied immunohistochemistry & molecular morphology**, Hagerstown, v. 17, n. 4, p. 319-28, jul. 2009.
- NOUROZ, F.; NOREEN, S.; HESLOP-HARRISON, J. S. Identification and characterization of LTR Retrotransposons in *Brassica*. **Turkish Journal of Biology**, Ankara, v. 39, p. 740-757, mar. 2015.
- NOUROZ, F.; NOREEN, S.; HESLOP-HARRISON, J. S. Identification and evolutionary dynamics of cacta dna transposons in brassica. **Pakistan Journal of Botany**, Karachi, v. 49, n. 2, p. 789-798, jan. 2017.
- NOVAK, P.; NEUMANN, P.; MACAS, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. **Bio Med Central bioinformatics**, London, v. 11, p. 378, jul. 2010.
- NOVAK, P.; NEUMANN, P.; PECH, J.; STEINHAI SL, J.; MACAS, J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. **Bioinformatics**, Oxford, v. 29, p. 792-793, mar. 2013.
- NOWOTNY, M. Retroviral integrase superfamily: the structural perspective. **European Molecular Biology Organization reports**, Oxford, v. 10, n. 2, p. 144-151, feb. 2009.
- O'CONNEL, M.; WRIGTH, J. Microsatellite DNA in fishes. **Reviews in Fish Biology and Fisheries**, Townsville, v. 7, p. 331-363, sep. 1997.
- OHNO, S. So much "junk" DNA in our genome. In: SMITH, H. H. (Ed.). **Evolution of Genetic Systems**. New York: Gordon and Breach, 1972. p. 366-370.
- OKAZAKI, S.; ISHIKAWA, H.; FUJIWARA, H. Structural analysis of TRAS1, a novel family of telomeric repeat-associated retrotransposons in the silkworm, *Bombyx mori*. **Molecular and cellular biology**, Washington, v. 15, n. 8, p. 4545-4552, aug. 1995.
- OLIVEIRA, C.; WANG Y.; BRYDEN L. J.; WRIGHT, J. M. Short interspersed repetitive elements (SINEs) from the cichlid fish, *Oreochromis niloticus*, and their chromosomal localization by fluorescent in situ hybridization **CARYOLOGIA**, Firenze, v. 56, n. 2, p. 181-189, apr. 2003.
- OSTERTAG, E. M.; KAZAZIAN JUNIOR, H. H. Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. **Genome research**, New York, v.11, p. 2059-2065, dec. 2001.
- PALAZZO A. F.; GREGORY T. R. The Case for Junk DNA. **Public Library of Science genetics**, San Francisco, v. 10, n. 5, p. 1-8, may. 2014

- PAVANELLI, C. S. **Revisão taxonômica da família Parodontidae (Ostariophysi: Characiformes)**. 332 f. Tese (Doutorado em Zoologia). Setor de Ciências Biológicas. Universidade Federal de São Carlos, São Carlos, 1999.
- PAVANELLI, C. S.; BRITSKI, H. A. *Apareiodon Eigenmann*, 1916 (Teleostei, Characiformes), from the Tocantins-Araguaia Basin, with Description of three New species. **Copeia**, Lawrence, v. 2003, n. 2, p. 337-348, jun. 2003.
- PAVANELLI, C. S. Família Parodontidae In: BUCKUP, A. B.; MENEZES, N. A.; GHAZZI, M. S. (Ed.). **Catálogo das espécies de peixes de água doce do Brasil**. Rio de Janeiro: Museu Nacional, 2007. p. 18-19.
- PEREIRA, A.; CUYPERS, H.; GIERL, A.; SCHWARZ-SOMMER, Z.; SAEDLER, H. Molecular analysis of the En/Spm transposable element system of *Zea mays*. **The EMBO Journal**. Oxford, v. 5, n. 5, p. 835-841, feb. 1986.
- PEREIRA, A.; SAEDLER, H. Transpositional behavior of the maize En/Spm element in transgenic tobacco. **The EMBO journal**, Oxford, v. 8, p. 1315-1321, may. 1989.
- PEREIRA, C. S. A.; PAZIAN, M. F.; RÁB P.; COLLARES-PEREIRA M. J. Dynamics of Rex3 in the genomes of endangered *Iberian Leuciscinae* (Teleostei, Cyprinidae) and their natural hybrids. **Molecular cytogenetics**, London, v. 8, n. 81, p. 1-10, oct. 2015.
- PERMAL, E.; FLUTRE, T.; QUESNEVILLE, H. Roadmap for annotation transposable elements in eukaryote genomes. **Methods in molecular biology**, Clifton, v. 859, p. 53-68, jan. 2012.
- PETERSON-BURCH, B. D.; VOYTAS, D. F. Genes of the Pseudoviridae (Ty1/copia Retrotransposons). **Molecular biology and evolution**, Chicago, v. 19, n. 11, p. 1832-1845, nov. 2002.
- PETERSON, P. A. A Mutable Pale-Green Locus in Maize. **Genetics**, Illinois, v. 38, p. 682-683, feb, 1953.
- PINKEL, D.; STRAUME, T.; GRAY, J. W. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 83, p. 2934-2938, may. 1986.
- PLATT, R. N.; BLANCO-BERDUGO, L.; RAY, D. A. Accurate transposable element annotation is vital when analyzing new genome assemblies. **Genome biology and evolution**, Oxford, v. 8, p. 403-410, feb. 2016.
- PLASTERK, R. H.; IZSVÁK, Z.; IVICS, Z. Resident aliens: the Tc1/mariner superfamily of transposable elements. **Trends in genetics**, Amsterdam, v. 15, n. 8, p. 326-332, aug. 1999.

- PLOHL, M.; BORSTNIK, B.; LUCIJNICJUSTVI, A.; UGARKOVDI, N. D. Evidence for random distribution of sequence variants in *Tenebrio molitor* satellite DNA. **Genetical research**, London, v. 60, p. 7-13, aug. 1992.
- PLOHL, M.; LUCHETTI, A.; MEŠTROVIĆ, N.; MANTOVANI, B. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero) chromatin. **Gene**, Amsterdam, v. 409, p. 72-82, feb. 2008.
- PONS, J.; GILLESPIE, R. G. Common origin of the satellite DNAs of the Hawaiian spiders of the genus *Tetragnatha*: evolutionary constraints on the length and nucleotide composition of the repeats. **Gene**, Amstersam, v. 313, p. 169-177, aug. 2003.
- PONS, J.; JUAN, C.; PETITPIERRE, E. Higher-order organization and compartmentalization of satellite DNA PIM357 in species of the coleopteran genus *Pimelia*. **Chromosome Research**, Oxford, v. 10, n. 7, p. 597-606, feb. 2002.
- PRIMO, C. B.; GLUGOSKI, L.; VICARI, M. R.; NOGAROTO, V. Chromosome mapping and molecular characterization of the Tc1/Mariner Element in *Rineloricaria* (Siluriformes: Loricariidae). **Brazilian Archives of Biology and Technology**, Curitiba, v. 61, oct. 2018.
- PUCCI, M. B.; BARBOSA, P.; NOGAROTO, V.; ALMEIDA, M. C.; ARTONI, R. F.; SCACCHETTI, P. C.; PANSONATO-ALVES, J. C.; FORESTI, F.; MOREIRA-FILHO, O.; VICARI, M. R. Chromosomal spreading of microsatellites and (TTAGGG)<sub>n</sub> sequences in the *Characidium zebra* and *C. gomesi* genomes (Characiformes: Crenuchidae). **Cytogenetic and Genome Research**, Würzburg, v. 149, p. 182-190, aug. 2016.
- PUCCI, M. B.; NOGAROTO, V.; MOREIRA-FILHO, O.; VICARI, M. R. Dispersion of transposable elements and multigene families: Microstructural variation in *Characidium* (Characiformes: Crenuchidae) genomes. **Genetics and Molecular Biology**, São Paulo, v. 41, n. 3, p. 585-592, 2018.
- PUIG, M.; CÁCERES, M.; RUIZ, A. Silencing of a gene adjacent to the breakpoint of a widespread *Drosophila* inversion by a transposon-induced antisense RNA. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 101, n. 24, p. 9013-9018, jun. 2004.
- RAMOS, K. S.; HE, Q.; KALBFLEISCH, T.; MONTOYA-DURANGO, D. E.; TENENG, I.; STRIBINSKIS, V.; BRUN, M. Computational and biological inference of gene regulatory networks of the LINE-1 retrotransposon. **Genomics**, London, v. 90, p. 176-185, aug. 2007.
- RAMSAY, L.; MACAULAY, M.; CARDLE, L.; MORGANTE, M.; IVANISSEVICH, S. D.; MAESTRI, E.; POWELL, W.; WAUGH, R. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. **The Plant Journal**, Medford, v. 17, n. 4, p. 415-425, mar. 1999.

- RAWLINGS, N. D.; MORTON, F. R.; KOK, C. Y.; KONG, J.; BARRETT, A. J. MEROPS: the peptidase database. **Nucleic Acids Research**, London, v. 36, p. 320-325, jan. 2008.
- RIBEIRO, L. B.; MORAES-NETO, A.; ARTONI, R. F.; MATOSO, D. A.; FELDBERG, E. Chromosomal Mapping of Repetitive Sequences (Rex3, Rex6, and rDNA Genes) in Hybrids Between *Colossoma macropomum* (Cuvier, 1818) and *Piaractus mesopotamicus* (Holmberg, 1887). **Zebrafish**, New York, v. 14, n. 2, p. 155-160, apr. 2017.
- RICHARD, G. F.; DUJON, B. Molecular evolution of minisatellites in *hemiascomycetous yeasts*. **Molecular Biology and Evolution**, v. 23, n. 1, p. 189-202, jan. 2006.
- RIO, D. C. P Transposable Elements in *Drosophila melanogaster*. In: CRAIG, N.; CRAIGIE, R.; GELLERT, M.; LAMBOWITZ, A. (Ed.). **Mobile DNA II**. Washington: ASM Press, 2002. p. 484-518.
- ROBERTS, T. R. Dental polymorphism and systematics in *Saccodon*, a neotropical genus of freshwater fishes (Parodontidae, Characoidei). **Journal of Zoology**, Medford, v. 173, n. 3, p. 303-321, jul. 1974.
- ROEST-CROLLIUS, H.; JAILLON, O.; DASILVA, C.; OZOUF-COSTAZ, C.; FIZAMES, C.; FISCHER, C.; BOUNEAU, L.; BILLAULT, A.; QUETIER, F.; SAURIN, W.; BERNOT, A.; WEISSENBAACH, J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. **Genome Research**, New York, v. 10, p. 939-949, may. 2000.
- ROHDE, C. Polimorfismo cromossômico e elementos transponíveis em *Drosophila willistoni*. 238 f. Tese (Doutorado em Genética e Biologia Molecular), Setor de Ciências Biológicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.
- ROSA, R.; BELLAFRONTE, E.; MOREIRA-FILHO, O.; MARGARIDO, V. P. Description of the ZZ/ZW sex chromosome system and localization of 5S and 18S rDNA genes in *Apareiodon* sp. (Pisces, Characiformes, Parodontidae). **Genetica**, s'-Gravenhage, v. 128, p. 159-166, feb. 2006.
- ROSA, R. S.; LIMA, F. C. T. Os peixes brasileiros ameaçados de extinção. In: MACHADO, A. B. M.; DRUMMOND, G. M.; PAGLIA, A. P. **Livro vermelho da fauna brasileira ameaçada de extinção**. 2 ed. Ministério do Meio Ambiente, Brasília, 2008. p. 9-19.
- ROSENFELD, M. R.; EICHEN, J. G.; WADE, D. F.; POSNER, J. B.; DALMAU, J. Molecular and clinical diversity in paraneoplastic immunity to Ma proteins. **Annals of Neurology**, Boston, v. 50, n. 3, p. 339-348, sep. 2001.
- ROZEN, S.; SKALETISKY, H. Primer3 on the WWW for general users and for biologist programmers. **Bioinformatics Methods and Protocols**, Clifton, v. 132, p. 365-386, feb. 2000.



- SABOT, F.; SCHULMAN, A. H. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. **Heredity**, Basingstoke, v. 97, n. 6, p. 381-388, dec. 2006.
- SAIGO, K.; KUGIMIYA, W.; MATSUO, Y.; INOUE, S.; YUKI, S. Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. **Nature**, London, v. 312, p. 659-661, dec. 1984.
- SAITOH, Y.; SAITOH, H.; OHTOMO, K.; MIZUNO, S. Occupancy of the majority of DNA in the chicken W chromosome by bent-repetitive sequences. **Chromosoma**, Berlin, v. 101, n. 1, p. 32-40, oct. 1991.
- SANTOS, E. O.; DEON, G. A.; ALMEIDA, R. B.; OLIVEIRA, E. A.; NOGAROTO, V.; SILVA, H. P.; PAVANELLI, C. S.; CESTARI, M. M.; BERTOLLO, L. A. C.; MOREIRA-FILHO, O.; VICARI, M. R. Cytogenetics and DNA barcode reveal an undescribed *Apareiodon* species (Characiformes: Parodontidae). **Genetics and Molecular Biology**. São Paulo. Em fase de pré-publicação.
- SAZIMA, I. Behavior of two brazilian species of parodontid fishes, *Apareiodon piracicabae* and *A. ibitiensis*. **Copeia**, v.1, p. 166-169, feb. 1980.
- SCACCHETTI, P. C.; UTSUNOMIA, R.; PANSONATO-ALVES, J. C.; DA COSTA SILVA, G. J.; VICARI, M. R.; ARTONI, R. F.; OLIVEIRA, C.; FORESTI, F. Repetitive DNA Sequences and Evolution of ZZ/ZW Sex Chromosomes in *Characidium* (Teleostei: Characiformes). **Public Library of Science One**, v. 10, n. 9, p. e0137231, sep. 2015.
- SCHEMBERGER, M. O.; BELLAFRONTE, E.; NOGAROTO, V.; ALMEIDA, M. C.; SCHÜHLI, G. S.; ARTONI, R. F.; VICARI, M. R. Differentiation of repetitive DNA sites and sex chromosome systems reveal closely related group in Parodontidae (Actinopterygii: Characiformes). **Genetica**, v. 139, n. 11-12, p.1499–1508, dec. 2011.
- SCHEMBERGER, M. O.; NASCIMENTO, V. D.; COAN, R.; RAMOS, E.; NOGAROTO, V.; ZIEMNICZAK, K.; VALENTE, G. T.; MARTINS, C.; MOREIRA-FILHO, O.; VICARI, M. R. DNA transposons invasion and microsatellites accumulation guide the W chromosome differentiation in a Neotropical fish genome. Submetido à publicação.
- SCHEMBERGER, M. O.; NOGAROTO, V.; ALMEIDA, M. C.; ARTONI, R. F.; VALENTE, T. G.; MARTINS, C.; MOREIRA-FILHO, O.; CESTARI, M. M.; VICARI, M. R. Sequence analyses and chromosomal distribution of the Tc1/Mariner element in Parodontidae fish (Teleostei: Characiformes). **Gene**, Amstersam, v. 593, n. 2, p. 308-314, nov. 2016.
- SCHEMBERGER, M. O.; OLIVEIRA, J. I. N.; NOGAROTO, V.; ALMEIDA, M. C.; ARTONI, R. F.; CESTARI, M. M.; MOREIRA-FILHO, O.; VICARI, M. R. Construction and characterization of a repetitive DNA library in Parodontidae (Actinopterygii: Characiformes): a genomic and evolutionary approach to the



- degeneration of the W sex chromosome. **Zebrafish**, New York, v. 11, n. 6, p. 518-527, dec. 2014.
- SCHWARZ-SOMMER, Z. S.; GIERL, A.; BERNDTGEN, R.; SAEDLER, H. Sequence comparison of 'states' of a1-m1 suggests a model of Spm (En) action. **EMBO Journal**, Oxford, v. 4, n. 10, p. 2439-2443, oct. 1985.
- SENE, V. F.; PANSONATO-ALVES, J. C.; FERREIRA, D. C.; UTSUNOMIA, R.; OLIVEIRA, C.; FORESTI, F. Mapping of the Retrotransposable Elements Rex1 and Rex3 in Chromosomes of Eigenmannia (Teleostei, Gymnotiformes, Sternopygidae). **Cytogenetic and Genome Research**, Würzburg, v. 146, n. 6, p. 319-324, nov. 2015.
- SHAO, F.; WANG, J.; XU, H.; PENG, Z. FishTEDB: a collective database of transposable elements identified in the complete genomes of fish. **Database**, Oxford, v. 2018, p. 1-9, jan. 2018.
- SHEHU-XHILAGA. M.; CROWE, S.M.; MAK, J. Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. **Journal of Virology**, Baltimore, v. 75, n. 4, p. 1834-1841, feb. 2001.
- SHEN, C. H.; STEINER, L. A. Genome structure and thymic expression of an endogenous retrovirus in zebrafish. **Journal of Virology**, Baltimore, v. 78, n. 2, p. 899-911, jan. 2004.
- SINZELLE, L.; IZSVÁK, Z.; IVICS, Z. Molecular domestication of transposable elements: From detrimental parasites to useful host genes. **Cellular and Molecular Life Sciences**, Basel, v. 66, n. 6, p. 1073-1093, mar. 2009.
- SLOTKIN, R. K. The case for not masking away repetitive DNA. **Mobile DNA**, London, v. 9, n. 15, p. 1-4, may. 2018.
- SLOTKIN, R. K.; MARTIENSSEN, R. Transposable elements and the epigenetic regulation of the genome. **Nature Reviews Genetics**, London, v. 8, n. 4, p. 272-285, apr. 2007.
- SMIT, A. F. A.; HUBLEY, R. RepeatModeler. Open-1.0 (2008-2015). Disponível em: <<http://www.repeatmasker.org>>. Acesso em: 26 nov. 2017.
- SMIT, A. F. A.; HUBLEY, R.; GREEN, P. RepeatMasker Open-4.0. 2013-2015. Disponível em: <<http://www.repeatmasker.org>>. Acesso em: 20 set. 2017.
- SMITH, G. P. Unequal crossover and evolution of multigene families. **Cold Spring Harbor Symposia on Quantitative Biology**, New York, v. 38, p. 507-513, jan. 1974.
- SNUSTAD, P.; SIMMONS, M. J. **Fundamentos de genética**. 6. ed. Rio de Janeiro: Guanabara Koogan, 2013.

- SPERBER, G. O.; AIROLA, T.; JERN, P.; BLOMBERG, J. Automated recognition of retroviral sequences in genomic data-RetroTector. **Nucleic Acids Research**, London, v. 35, n. 15, p. 1-13, aug. 2007.
- SOBIESKI, D. A.; EDEN, F. C. Clustering and methylation of repeated DNA: persistence in avian development and evolution. **Nucleic Acids Research**, London, v. 9, n. 22, p. 6001-6015, nov. 1981.
- SOUSA, S. M. de, CARNEIRO, A. A., CARNEIRO, N. P. Decifrando o genoma em grande escala. **Informe Agropecuário**, Belo Horizonte, v. 30, p. 108-115, nov./dec. 2009.
- STEINEMANN, M.; STEINEMANN, S. Enigma of Y chromosome degeneration: Neo-Y and Neo-X chromosomes of *Drosophila miranda* a model for sex chromosome evolution. **Genetica**, v. 102-103, n. 1-6, p. 409-420, feb, 1998.
- TAUTZ, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. **Nucleic Acids Research**, London, v. 17, n. 16, p. 6463-6471, aug. 1989.
- TELENTI, A.; MARTINEZ, R.; MUNOZ, M.; BLEIBER, G.; GREUB, G.; SANGLARD, D.; PETERS, S. Analysis of natural variants of the human immunodeficiency virus type 1 gag-pol frameshift stem loop structure. **Journal of Virology**, Baltimore, v. 76, n. 15, p. 7868-7873, aug. 2002.
- THE GENOME INTERNATIONAL SEQUENCING CONSORTIUM. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860-921, feb. 2001.
- TACKE, E.; SCHWARZ-SOMMER, Z. S.; PETERSON, P. A.. SAEDLER, H. Molecular Analysis of states of A1 locus of *Zea Mays*. **Maydica**, Bégamo, v. 31, p. 83-91, fev. 1986.
- TING, C. N.; ROSENBERG, M. P.; SNOW, C. M.; SAMUELSON, L. C.; MEISLER, M. H. Endogenousretroviral sequences are required for tissue-specific expression of a human salivary amylase gene. **Genes & Development**, v. 6, n. 8, p. 1457-1465, aug. 1992.
- TRALDI, J. B.; VICARI, M. R.; MARTINEZ, J. F.; BLANCO, D. R.; LUI, R. L.; MOREIRA-FILHO, O. Chromosome Analyses of *Apareiodon argenteus* and *Apareiodon davisii* (Characiformes, Parodontidae): An Extensive Chromosomal Polymorphism of 45S and 5S Ribosomal DNAs. **Zebrafish**, New York, v. 13, n. 1, p. 19-25, feb. 2016.
- TRAVASSOS, H. Ictiofana de Pirassununga. I- Subfamília Parodontidae Eigenmann, 1916 (Actinopterygii-Cypriniformes). **Boletim do museu nacional**, Rio de Janeiro, v. 129, p. 1-31, 1955.
- TRENTMANN, S. M.; SAEDLER, H.; GIERL, A. The transposable element *En/Spm*-encoded TNPA protein contains a DNA binding and a dimerization domain. **Molecular and General Genetics**, v. 238, p. 201-208, apr. 1993.

UGARKOVIC, D.; PLOHL, M. Variation in satellite DNA profiles causes and effects. **The EMBO Journal**, Oxford, v. 21, n. 22, p. 5955-5959, nov. 2002.

UNTERGASSER, A.; NIJVEEN, H. Primer 3Plus. 2007. Disponível em:  
<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>

VALENTE, G. T.; CONTE, M. A.; FANTINATTI, B. E.; CABRAL-DE-MELLO, D. C.; CARVALHO, R. F.; VICARI, M. R.; KOCHER, T. D.; MARTINS, C. Origin and evolution of B chromosomes in the cichlid fish *Astatotilapia latifasciata* based on integrated genomic analyses. **Molecular Biology and Evolution**, Oxford, v. 31, n. 8, p. 2061-2072, aug. 2014.

VALENTE, G. T.; NAKAJIMA, R. T.; FANTINATTI, B. E. A.; MARQUES, D. F.; ALMEIDA, R. O.; SIMÕES, R. P.; MARTINS, C. B chromosomes: from cytogenetics to systems biology. **Chromosoma**, Wien, v. 126, n. 1, p. 73-81, feb. 2017.

VERGNAUD, G.; DENOEUDE, F. Minisatellites: mutability and genome architecture. **Genome Research**, New York, v. 10, n. 7, p. 899-907, jul. 2000.

VERNHETTES, S.; GRANDBASTIEN, M. A.; CASACUBERTA, J. M. In vivo characterization of transcriptional regulatory sequences involved in the defence-associated expression of the tobacco retrotransposon Tnt1. **Plant Molecular Biology**, v. 35, n. 5, p. 673-679, nov. 1997.

VICARI, M. R.; MOREIRA-FILHO, O.; ARTONI, R. F.; BERTOLLO, L. A. C. ZZ/ZW sex chromosome system in an undescribed species of the genus *Apareiodon* (Characiformes, Parodontidae). **Cytogenetic and Genome Research**, Würzburg, v. 114, p. 163-168, 2006.

VICARI, M. R.; NOGAROTO, V.; NOLETO, R. B.; CESTARI, M. M.; CIOFFI, M. B.; ALMEIDA, M. C.; MOREIRA-FILHO, O.; BERTOLLO, L. A. C.; ARTONI, R. F. Satellite DNA and chromosomes in neotropical fishes: methods, applications and perspectives. **Journal of Fish Biology**, Liverpool, v. 76, n. 5, p. 1094-1116, apr. 2010.

VICENTE, V. E.; BERTOLLO, L. A. C.; VALENTINI, S. R.; MOREIRA-FILHO, O. Origin and differentiation of sex chromosome system in *Parodon hilarii* (Pisces, Parodontidae). Satellite DNA, G and C-banding. **Genetica**, v. 119, n. 2, p. 115-120, oct. 2003.

VICENTE, V. E.; JESUS, C. M.; MOREIRA-FILHO, O. Chromosomal localization of 5S and 18S rDNA genes in three *Parodon* species (Pisces, Parodontidae). **Caryologia**, Firenze, v. 54, n. 4, p. 365-369, oct. 2001.

VIEIRA, C.; NARDON, C.; ARPIN, C.; LEPETIT, D.; BIÉMONT, C. Evolution of genome size in *Drosophila*. Is the Invader's Genome Being Invaded by Transposable Elements? **Molecular Biology and Evolution**, v. 19, n. 7, p. 1154-1161, jul. 2002.

- VOLFF, J. N. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. **BioEssays**, Cambridge, v. 28, n. 9, p. 913–922, sep. 2006.
- VOLFF, J. N. Cellular Genes Derived from Gypsy/Ty3 Retrotransposons in Mammalian Genomes. **Annals of the New York Academy of Science**, New York, v. 1178, p. 233-243, oct. 2009.
- VOLFF, J. N.; KÖRTING, C.; FROSCHAUER, A.; SWEENEY, K.; SCHARTL, M. Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates. **Journal of Molecular Evolution**, New York, v. 52, n. 4, p. 351-360, apr. 2001.
- WAGNER, R. P.; MAGUIRE, M. P.; STALLINGS, R. L. Chromosomes: A synthesis. New York: Wiley-Liss, 1993.
- WAN, Q. H.; WU, H.; FUJIHARA, T.; FANG, S. G. Which genetic marker for which conservation genetics issue?. **Electrophoresis**, Weinheim, v. 25, p. 2165-2176, jul. 2004.
- WARREN, I. A.; NAVILLE, M.; CHALOPIN, D.; LEVIN, P.; BERGER, C. S.; GALIANA, D.; VOLFF, J. N. Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. **Chromosome Research**, Oxford, v. 23, n. 3, p. 505-531, sep. 2015.
- WEBER, J. L.; MAY, P.E. Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. **The American Journal of Human Genetics**, Baltimore, v. 44, n. 3, p. 388-396, mar. 1989.
- WICKER, T.; GUNDLACH, H.; SPANNAGL, M.; UAUY, C.; BORRILL, P.; RAMÍREZ-GONZÁLEZ, R. H.; OLIVEIRA, R.; International Wheat Genome Sequencing Consortium; MAYER, K. F. X.; PAUX, E.; CHOULET, F. Impact of transposable elements on genome structure and evolution in bread wheat. **Genome Biology**, London, v. 19, n. 103, p. 1-18, aug. 2018.
- WICKER, T.; SABOT, F.; HUA-VAN, A.; BENNETZEN, J. L.; CAPY, P.; CHALHOUB, B.; FLAVELL, A.; LEROY, P.; MORGANTE, M.; PANAUD, O.; PAUX, E.; SANMIGUEL, P.; SCHULMAN, A H. A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**, London, v. 8, n. 12, p. 973-982, nov. 2007.
- WILDER, J.; HOLLOCHER, H. Mobile elements and the genesis of microsatellites in dipterans. **Molecular Biology and Evolution**, Oxford, v. 18, n. 3, p. 384-392, mar. 2001.
- WILLIAMS, A. J.; BLACKLOW, S. C.; COLLINS, T. The Zinc Finger-Associated SCAN Box Is a Conserved Oligomerization Domain. **Molecular and Cellular Biology**, Washington, v. 19, n. 12, p. 8526-8535, dec. 1999.

- XIONG, Y.; EICKBUSH, T. H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. **The EMBO Journal**, Oxford, v. 9, n. 10, p. 3353-3362, oct. 1990.
- YE, C.; JI, G.; LI, L.; LIANG, C. detectIR: A Novel Program for Detecting Perfect and Imperfect Inverted Repeats Using Complex Numbers and Vector Calculation. **Public Library of Science One**, San Francisco, v. 9, n. 11, p. e113349, nov. 2014.
- YUAN, Z.; LIU, S.; ZHOU, T.; TIAN, C.; BAO, L.; DUNHAM, R.; LIU, Z. Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. **BioMed Central Genomics**, London, v. 19, n. 141, p. 1-10, feb. 2018.
- ZELENTSOVA, E. S.; VASHAKIDZE, R. P.; KRAYEV, A. S.; EVGEN'EV, M. B. Dispersed repeats in *Drosophila virilis*: Elements mobilized by interspecific hybridization. **Chromosoma**, Wien, v. 93, n. 6, p. 469-476, jun. 1986.
- ZERBINO, D. R.; BIRNEY, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. **Genome Research**, New York, v. 18, n. 5, p. 821-829, may. 2008.
- ZHANG, H. H.; LI, G. Y.; XIONG, X. M.; HAN, M. J.; ZHANG, X. G.; DAI, F. Y. TRT, a Vertebrate and Protozoan Tc1-Like Transposon: Current Activity and Horizontal Transfer. **Genome Biology and Evolution**, Oxford, v. 8, n. 9, p. 2994-3005, oct. 2016.
- ZHOU, Q. A swimy locus on Y chromosome of the platyfish (*Xiphophorus maculatus*) is derived from a novel DNA transposon Zisupton. **Gene**, Amsterdam, v. 503, n. 2, p.254-259, jul. 2012.
- ZIEMNICZAK, K.; TRALDI, J. B.; NOGAROTO, V.; ALMEIDA, M. C.; ARTONI, R. F.; MOREIRA-FILHO, O.; VICARI, M. R. In situ localization of (GATA)<sub>n</sub> and (TTAGGG)<sub>n</sub> repeated DNAs and W sex chromosome differentiation in Parodontidae (Actinopterygii: Characiformes). **Cytogenetic and Genome Research**, Würzburg, v. 144, n. 4, p. 325-332, fev. 2014.



# ANEXO 1. Licença permanente para coleta de material zoológico MMA/IBAMA/SISBIO.



Ministério do Meio Ambiente - MMA  
Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis - IBAMA  
Instituto Chico Mendes de Conservação da Biodiversidade - ICMBio  
Sistema de Autorização e Informação em Biodiversidade - SISBIO

## Licença permanente para coleta de material zoológico

Número: 15117-1		Data da Emissão: 26/03/2008 18:25
Dados do titular		
Registro no Ibama: 2537361	Nome: Marcelo Ricardo Vicari	CPF: 952.846.480-72
Nome da Instituição: UNIVERSIDADE ESTADUAL DE PONTA GROSSA		CNPJ: 80.257.355/0001-08

### Observações, ressalvas e condicionantes

1	A participação de pesquisador(a) estrangeiro(a) nas atividades previstas nesta autorização depende de autorização expedida pelo Ministério de Ciência e Tecnologia (CNPq/MCT).
2	A licença permanente não é válida para: a) coleta ou transporte de espécies que constem nas listas oficiais de espécies ameaçadas de extinção; b) manutenção de espécimes de fauna silvestre em cativeiro; c) recebimento ou envio de material biológico ao exterior; e d) realização de pesquisa em unidade de conservação federal ou em caverna. A restrição prevista no item d não se aplica às categorias Reserva Particular do Patrimônio Natural, Área de Relevante Interesse Ecológico e Área de Proteção Ambiental constituídas por terras privadas.
3	O pesquisador titular da licença permanente, quando acompanhado, deverá registrar a expedição de campo no Sisbio e informar o nome e CPF dos membros da sua equipe, bem como dados da expedição, que constarão no comprovante de registro de expedição para eventual apresentação à fiscalização;
4	Esta licença permanente não exime o seu titular da necessidade de obter as anuências previstas em outros instrumentos legais, bem como do consentimento do responsável pela área, pública ou privada, onde será realizada a atividade.
5	Esta licença permanente não poderá ser utilizada para fins comerciais, industriais, esportivos ou para realização de atividades inerentes ao processo de licenciamento ambiental de empreendimentos.
6	Esta autorização NÃO exime o pesquisador titular da necessidade de atender ao disposto na Instrução Normativa Ibama nº 27/2002, que regulamenta o Sistema Nacional de Anilhamento de Aves Silvestres.
7	O pesquisador titular da licença permanente será responsável pelos atos dos membros da equipe (quando for o caso).
8	O órgão gestor de unidade de conservação estadual, distrital ou municipal poderá, a despeito da licença permanente e das autorizações concedidas pelo Ibama, estabelecer outras condições para a realização de pesquisa nessas unidades de conservação.
9	O titular de licença ou autorização e os membros da sua equipe deverão optar por métodos de coleta e instrumentos de captura direcionados, sempre que possível, ao grupo taxonômico de interesse, evitando a morte ou dano significativo a outros grupos; e empregar esforço de coleta ou captura que não comprometa a viabilidade de populações do grupo taxonômico de interesse em condições in situ.
10	O titular da licença permanente deverá apresentar, anualmente, relatório de atividades a ser enviado por meio do Sisbio no prazo de até 30 dias após o aniversário de emissão da licença permanente.
11	O titular de autorização ou de licença permanente, assim como os membros de sua equipe, quando da violação da legislação vigente, ou quando da inadequação, omissão ou falsa descrição de informações relevantes que subsidiaram a expedição do ato, poderá, mediante decisão motivada, ter a autorização ou licença suspensa ou revogada pelo Ibama e o material biológico coletado apreendido nos termos da legislação brasileira em vigor.
12	A licença permanente será válida enquanto durar o vínculo empregatício do pesquisador com a instituição científica a qual ele estava vinculado por ocasião da solicitação.
13	Este documento não dispensa o cumprimento da legislação que dispõe sobre acesso a componente do patrimônio genético existente no território nacional, na plataforma continental e na zona econômica exclusiva, ou ao conhecimento tradicional associado ao patrimônio genético, para fins de pesquisa científica, bioprospecção e desenvolvimento tecnológico.
14	As atividades contempladas nesta autorização NÃO abrangem espécies brasileiras constantes de listas oficiais (de abrangência nacional, estadual ou municipal) de espécies ameaçadas de extinção, sobreexplotadas ou ameaçadas de sobreexploração.

### Táxons autorizados

#	Nível taxonômico	Táxon(s)
1	ORDEM	Characiformes, Cypriniformes, Synbranchiformes, Perciformes, Siluriformes, Gymnotiformes

### Destino do material biológico coletado

#	Nome local destino	Tipo Destino
1	UNIVERSIDADE ESTADUAL DE PONTA GROSSA	Citogenética de Peixes
2	Museu Nacional da Universidade Federal do Rio de Janeiro	coleta

Este documento (Licença permanente para coleta de material zoológico) foi expedido com base na Instrução Normativa Ibama nº154/2007. Através do código de autenticação abaixo, qualquer cidadão poderá verificar a autenticidade ou regularidade deste documento, por meio da página do Ibama/Sisbio na Internet ([www.ibama.gov.br/sisbio](http://www.ibama.gov.br/sisbio)).

Código de autenticação: 61434795



Página 1/2



**ANEXO 2.** Carta de aprovação pelo Comitê de Ética em Experimentação Animal da Universidade Estadual de Ponta Grossa.



UNIVERSIDADE ESTADUAL DE PONTA GROSSA

**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
COMISSÃO DE ÉTICA DO USO DE ANIMAL**

**CARTA DE APROVAÇÃO**

**Processo CEUA – 011/2017**

**Protocolo UEPG – 6130/2017**

**Título** – Solicitação de utilização de animais (peixes) no projeto intitulado: Citogenética molecular em parodontidae loricariidae: caracterização de genes e DNAs repetitivos.  
**Interessado** – Prof. Marcelo Ricardo Vicari

**Data de Entrada** – 5/4/2017

**Resultado:** Aprovado

**Data/Prazo** – 28/04/2017 a 28/04/2019

**Considerações**

Professor Marcelo Vicari (DEBIOGEM, vicarimr@yahoo.com.br)

Em relação a utilização de animais no protocolo de pesquisa sob sua responsabilidade, a CEUA deliberou pela sua aprovação pela aprovação, por dois anos, do uso de 60 (sessenta) peixes de peso variável.

Ponta Grossa, 03 de maio de 2017

UNIVERSIDADE ESTADUAL DE PONTA GROSSA  
Pró-Reitoria de Pesquisa e Pós-Graduação  
CEUA – Comissão de Ética e Pesquisas em Animais  
  
Dra. Dionízia Xavier Scamporrin  
Coordenadora

Av. Gen. Carlos Cavalcanti, nº 4748, CEP 84.030-900 Campus Universitário em Uvaranas  
Ponta Grossa – Paraná  
Bloco da Reitoria – anexo a PROPEP  
Fone: (042) 3220-3264

**ANEXO 3.** Conjuntos de primers sintetizados para validação RTE Gypsy e um conjunto para o transposon En/Spm.

	IDENTIFICAÇÃO	SEQUÊNCIA
<b>Conjunto 1</b>	Gypsy_P1_Forward	5' GCTAGCAAGTGGTCTCTATGAA 3'
	Gypsy_P1_Reverse	5' AGGCACAGATAGGGCTATAGAA 3'
<b>Conjunto 2</b>	Gypsy_P2_Forward	5' CCCCACAGCTCTTTATGTAGT 3'
	Gypsy_P2_Reverse	5' TGGAGAGAGGTGGACTGAA 3'
<b>Conjunto 3</b>	Gypsy_P3_Forward	5' GGTCAGCAAAAGCCAGTAG 3'
	Gypsy_P3_Reverse	5' CACATCAGGCTCAAGGACT 3'
<b>Conjunto 4</b>	Gypsy_P4_Forward	5' CAGCTAGCAGCTCCTGAA 3'
	Gypsy_P4_Reverse	5' GGATAGCGGGTCTCAAGT 3'
<b>Conjunto 5</b>	Gypsy_P5_Forward	5' TGAATGAACCCTTTATTGTCA 3'
	Gypsy_P5_Reverse	5' CTAAGGAGGTTGGAGCAAG 5'
<b>Conjunto 6</b>	Gypsy_P6_Forward	5' GTTACTGCACCTCTCAAAAGC 3'
	Gypsy_P6_Reverse	5' GGGGAGGGCCTAAGAGTA 5'
<b>Conjunto 7</b>	Enspm_Fw	5' GCAGGCTTTCACGAATCT 3'
	Enspm_Rv	5' CATGAGGGCTCCTGTTCT 3'

FONTE: O Autor (2018).

**ANEXO 4.** Extração de DNA - protocolo adaptado de Doyle e Doyle (1990).

A extração do DNA genômico e mitocondrial foi realizada a partir de tecidos sólidos (fígado ou músculo). Para isto, foi retirado um pequeno pedaço (15 mm<sup>3</sup>) do fígado ou músculo do animal e este foi colocado em um microtubo contendo 360 µl de tampão (NaCl 5 M, EDTA 0,5 M, TRIS-HCl 1 M) sem CTAB (brometo de cetil-trimetilamônio). Após, foi acrescentado 13 µl de proteinase K (10 mg/mL), 5 µl de β-mercaptoetanol e 240 µl de CTAB 5% cristalizado. Posteriormente a solução foi incubada em banho-maria a 55°C, *overnight* ou, até digerir totalmente o tecido. Após este período foram adicionados 600 µl de clorofórmio sob agitação por 5 minutos em vortex suave. Em seguida, o material foi centrifugado a 8000 rpm por 5 minutos e o sobrenadante foi transferido para um novo tubo. Neste foram adicionados 600 µl de etanol absoluto e reservado a – 20 °C por 2 horas ou *overnight*. Posteriormente, o material foi centrifugado a 14.000 rpm por 15 minutos e o sobrenadante descartado. O *pellet* foi lavado com 500 µl de etanol a 70 v/v, sem agitar. A seguir, o material foi centrifugado por 2 minutos a 14.000 rpm e o sobrenadante descartado. O *pellet* foi seco em estufa a 37°C. Posteriormente, foram adicionados 5 µl de RNase 20 mg/mL e o material foi mantido na estufa a 37 °C por 1 h. Em seguida foi inativado a RNase a 65 °C por 15 minutos e o DNA resultante foi armazenado no freezer.

**ANEXO 5.** Parâmetros utilizados para amplificação dos elementos Gypsy e En/Spm.

Reação de Amplificação			Long PCR Gypsy	PCR Gypsy	PCR En/Spm
	DNA genômico		200 ng	40ng	40 ng
	Primer Forward		0,2 µM	0,2 µM	0,2 µM
	Primers Reverse		0,2 µM	0,2 µM	0,2 µM
	Desoxinucleotídeos trifosfatos (dNTPs)		0,4 mM	0,4 mM	0,4 mM
	Tampão de reação (200 mM Tris, pH 8,4, 500 mM KCl)		1×	1×	1×
	MgCl <sub>2</sub>		2 mM	2 mM	2 mM
	Taq DNA polimerase		5 U Long Amplification	1 U	1 U
Parâmetros de amplificação	Desnaturação inicial		2 min a 95 °C	5 min a 94 °C	5 min a 94 °C
	Desnaturação	35 ciclos	45 seg. a 94 °C	1 min a 94 °C	1 min a 94 °C
	Anelamento dos primers		45 seg. a 55°C	45 seg. a 55°C	40 seg. a 56°C
	Extensão		7 min a 68 °C	2 min a 72 °C	45 seg. a 72 °C
	Extensão final		5 min a 68 °C	7 min a 72 °C	7 min a 72 °C
	Fragmento esperado		~5.600 pb	~1.000pb	510pb

FONTE: O autor (2019)

## **ANEXO 6.** Purificação do produto da PCR pelo kit da GE healthcare - illustraGFX PCR DNA

O protocolo consistiu na adição de 500 µl do capture buffer type3 à 100 µl de amostra (10 µg) e centrifugação por 30 segundos a 16.000 rcf. Posteriormente, o líquido do fundo do tubo que contém a amostra foi coletado e transferido para a microSpin Collumn. Esta foi centrifugada por 30 segundos a 16.000 rcf. Na sequência, todo o conteúdo do tubo coletor foi descartado, mantendo a microSpin Collumn. Em seguida, 500 µl da Wash buffer type1 foi adicionada a amostra e centrifugada por 30 segundos a 16.000 rcf. O conteúdo do tubo coletor foi descartado e novamente centrifugado por 20 segundos a 16.000 rcf. Posteriormente, a microSpin Collumn foi transferida para um microtubo novo, onde permaneceu para evaporar o álcool a 70° C por 3 minutos. Após, 25 µl do tampão type 6 a 70°C, foi adicionado a amostra, reservado por 1 minuto e, centrifugado por 1 minuto a 16.000 rcf. A solução que passou pela a microSpin Collumn foi coletada e estocada a – 20 °C.

**ANEXO 7.** Marcação de sondas por nick translation, kit dig-nick translation mix  
(Roche Applied Science)

As reações de nick translation seguiram as orientações do fabricante. Pipetar os seguintes componentes em tubo de 0,2 ml no gelo: x µl produto de PCR (equivalente a concentração de 1000-1500 ng/µl); 4 µL mix de digoxigenin nick translation; completar para o volume total de 20 µl com x µl água qsp; permanecer durante 1 hora e 30 minutos a 15 °C; interromper aquecendo por 15 minutos a 65 °C.



**ANEXO 8.** Suspensão de Levedura (CAVALLINI; BERTOLLO, 1988).

A técnica de suspensão de levedura consiste em uma mistura (0,5 g de fermento biológico + 1,5 g de dextrose + 6 mL de H<sub>2</sub>O) que deve ser injetada no peixe na região intraperitoneal em uma proporção de 1 mL / 100 g de peso animal. Este procedimento deve ser realizado 24 horas antes de iniciar as preparações de cromossomos mitóticos, com o intuito de induzir o aumento do número de células em divisão.

**ANEXO 9.** Preparações diretas para obtenção de cromossomos mitóticos  
(BERTOLLO; CIOFFI; MOREIRA-FILHO, 2015).

Na região intra-abdominal no animal injetou-se uma solução aquosa de colchicina 0,025%, na proporção de 1 ml/100 g de peso. Os peixes foram mantidos em aquário bem aerado durante 50-60 minutos, posteriormente foram anestesiados com benzocaína diluída a 0,01%, e sacrificados. Retirou-se a porção do rim anterior e dissociada com uma seringa desprovida de agulha em cerca de 10 ml de solução hipotônica (KCl 0,075 M). O tecido dissociado foi incubado em estufa a 37°C durante 25-30 minutos, depois re-suspendido com auxílio de uma pipeta Pasteur e transferido para tubo de centrifuga. Adicionou-se neste tubo 1ml de fixador (3 partes de metanol para 1 de ácido acético glacial), o material foi novamente re-suspendido várias vezes, centrifugado durante 10 minutos, a 900 rpm e depois descartado o sobrenadante. Após a centrifugação adicionou-se 5-7 ml do mesmo fixador, re-suspendido e centrifugado por mais 10 minutos a 900 rpm, este passo foi repetido 2 vezes. Descartou-se o sobrenadante e adicionou uma quantidade de fixador para que se obtivesse uma suspensão celular moderadamente concentrada (geralmente de 0,5 a 1,0 ml), re-suspendido e armazenado em freezer para posterior utilização.

## **ANEXO 10.** Hibridação in situ fluorescente (FISH) seguindo Pinkel et al. (1986).

### ***A. Preparação da solução de hibridação***

Em todas as FISH, a mistura de hibridação consistiu de: 25 µl formamida (50% de formamida); 10 µl sulfato de dextrano 50% (concentração final de 10%); 5 µl de 20x saline salt citrate - SSC (concentração final 2x SSC); 6 µl de H<sub>2</sub>O qsp, perfazendo um volume total de 400 µl, aos quais foram adicionados 4 µg da sonda (DNA marcado com digoxigenina 11dUTP). A estringência destas misturas de hibridação foi de 77%. Em seguida, a solução de hibridação foi levada ao termociclador, durante 10 minutos, a 99 °C para desnaturação do DNA e, imediatamente após, para um recipiente com gelo, impedindo a renaturação por choque térmico.

### ***B. Preparação das lâminas (lavagens pré-hibridação)***

As lâminas, contendo as preparações cromossômicas, foram lavadas em tampão fosfato (phosphate buffer saline-PBS), por 5 minutos, em temperatura ambiente e desidratadas em uma série de etanol a 70 v/v, 85 v/v e 95 v/v, 5 minutos em cada incubação. A seguir, foram tratadas com solução de RNase (100 µg/ml) durante 1 hora, em câmara úmida a 37°C, lavadas duas vezes em solução de 2xSSC, por 10 minutos e em PBS, por 5 minutos. Em seguida a fixação com formaldeído 1 %/ PBS 1x/MgCl<sub>2</sub> 50mM, por 10 minutos, à temperatura ambiente, lavagem em PBS 1x por 5 minutos e desidratação em série de etanol a 70 v/v, 85 v/v e 95 v/v, 5 minutos cada incubação. As lâminas foram então tratadas com formamida 70 % dissolvida em 2x SSC, a 70°C, por 5 minutos e novamente desidratadas em série de etanol gelado a 70 v/v, 85 v/v e 95 v/v, 5 minutos cada banho.

### ***C. Hibridação e detecção dos sinais correspondentes***

Foram aplicados, sobre as lâminas, cerca de 50 µl da solução de hibridação permanecendo “overnight” a 37°C, em câmara úmida contendo solução de 2x SSC pH 7,0. Decorrido este tempo, as lâminas foram lavadas com solução de 2x SSC pH 7,0 por 20 minutos. Em seguida foram lavadas em Tween 20, por 5 minutos, incubação em 90 µl de tampão non fat dry milk (NFDM) a 5 %, por 15 minutos em câmara úmida e duas lavagens com Tween 20, cinco minutos cada. Para a detecção das sondas, foram colocados sobre as lâminas uma mistura de 100 µl de tampão NFDM contendo 1:1000 de avidina-FITC (Fluoresceína Isotil Cianato-avidina

conjugada) a 0,25 µg/ml e; 5:1000 de anti digoxigenina rodamina a 1 µg/ml; permanecendo por 1 h a 37°C, em câmara úmida. As lâminas foram lavadas 3 vezes em Tween 20, cinco minutos cada. Em seguida a desidratação em série de etanol a 70 v/v, 85 v/v e 95 v/v, 5 minutos em cada banho. Os cromossomos foram contra corados com 4',6-diamidino-2-phenylindole (DAPI - 0,2 µg/ml) diluído em uma solução “antifade” (Fluka®).

**ANEXO 11.** Conjuntos de primers para análise de expressão do elemento Gypsy em *Apareiodon* sp.

IDENTIFICAÇÃO		SEQUÊNCIA	FRAGMENTO ESPERADO
<b>Integrase</b>	Forward	5' GGCCCTCATTTTGAAAGCC 3'	91pb
	Reverse	5' ACCCTGAGGATGATAAGCTGTG 3'	
<b>Reverse Transcriptase</b>	Forward	5' GCTCACTAATGCCCCAGCTAC 3'	175pb
	Reverse	5' CTTTAGTCCATCCTCCTGCAGTC 3'	
<b>PNMA</b>	Forward	5' AGCAGGGGAAAGAAGGCTCT 3'	129pb
	Reverse	5' ATTCCTAAACCAAGTGGTGAAGTTG 3'	

**ANEXO 12.** Protocolo de extração de RNA por TRIZOL (Invitrogen)

1. Picotar de 50-100 miligrama (mg) de tecido com uma lâmina de aço estéril;
2. Transferir para um tubo de 1,5 mililitro (ml);
3. Para cada 50-100 mg de tecido a ser processado, adicionar 1 ml de Trizol™ (a 15°C);
4. Agitar em vortex por 30 segundos;
5. Centrifugar a 12.000g RPM por 10 minutos a 4°C;
6. Transferir o sobrenadante para um novo tubo de 1,5ml
7. Incubar as amostras por 5 minutos em temperatura de 15-20°C;
8. Adicionar 200 microlitros (µl) de clorofórmio;
9. Agitar em vortex por 30 segundos;
10. Incubar as amostras por 15 minutos em temperatura de 15-20°C;
11. Centrifugar a 12.000g RPM por 15 minutos a 4°C;
12. Transferir o sobrenadante (parte incolor) para um novo tubo de 1,5ml;
13. Adicionar 500 µl de isopropanol;
14. Misturar os reagentes por inversão (3x);
15. Incubar as amostras por 10 minutos em temperatura de 15-20°C;
16. Centrifugar a 12.000g RPM por 10 minutos a 4°C;
17. Descartar o sobrenadante;
18. Adicionar 1ml de etanol a 70-75%;
19. Agitar em vortex por 30 segundos;
20. Centrifugar a 7.500g RPM por 5 minutos a 4°C;
21. Descartar o sobrenadante;
22. Repetir os passos 18 ao 21 por três vezes;
23. Adicionar 30 µl de H<sub>2</sub>O ultrapura e incubar por 20 minutos em temperatura de 15-20°C;
24. Aquecer em banho-maria de 55-60° por 10 min.



**ANEXO 13.** Protocolo de síntese de DNA complementar (cDNA) - Kit da GE  
Healthcare First-Strand cDNA Syntesis Kit

1. Em tubo de 0,2 ml adicionar X µl de RNA (1.000 nanogramas) e completar para 8 µl com RNase-free water;
2. Aquecer o RNA a 65°C por 10 minutos, passe imediatamente para o gelo;
3. Gentilmente adicional 5µl de bulk first-strand cDNA mix;
4. Adicionar 1 µl de DTT solution;
5. Pipetar 1 µl de primer random;
6. Incubar a 37°C por 1 hora;
7. Armazenar em freezer a -20°C.